Pradipta Maji Ashish Ghosh
M. Narasimha Murty Kuntal Ghosh
Sankar K. Pal (Eds.)

LNCS 8251

Pattern Recognition and Machine Intelligence

5th International Conference, PReMI 2013
Kolkata, India, December 2013
Proceedings

Lecture Notes in Computer Science 8251

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Pradipta Maji Ashish Ghosh
M. Narasimha Murty Kuntal Ghosh
Sankar K. Pal (Eds.)

Pattern Recognition and Machine Intelligence

5th International Conference, PReMI 2013
Kolkata, India, December 10-14, 2013
Proceedings

Springer

Volume Editors

Pradipta Maji
Ashish Ghosh
Kuntal Ghosh
Sankar K. Pal
Indian Statistical Institute
Machine Intelligence Unit
203, B. T. Road
Kolkata 700108, India
E-mail: {pmaji; ash; kuntal;
sankar}@isical.ac.in

M. Narasimha Murty
Indian Institute of Science
Department of Computer Science
and Automation
Bangalore 560012, India
E-mail: mnm@csa.iisc.ernet.in

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-45061-7 e-ISBN 978-3-642-45062-4
DOI 10.1007/978-3-642-45062-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013953502

CR Subject Classification (1998): I.4, F.1, I.2, I.5, J.3, H.3-4, K.4.4

LNCS Sublibrary:
SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

It is our pleasure to welcome you all to the 5th International Conference on Pattern Recognition and Machine Intelligence (PReMI 2013), being held in the Indian Statistical Institute, Kolkata, India, during December 10–14, 2013. The primary goal of the conference is to present state-of-the-art scientific results, encourage academic and industrial interaction, and promote collaborative research activities in pattern recognition, machine intelligence, and related fields, involving scientists, engineers, professionals, researchers, and students from India and abroad. The conference is held every two years to make it an ideal platform for people to share their views and experiences in the said areas. This is the fifth edition in this series.

The conference has one keynote speech, two plenary talks, and seven invited lecturers, all by very eminent and distinguished researchers from around the world. The conference had a very good response in terms of paper submissions. It received more than 300 submissions from about 22 countries spanning six continents. Each paper was critically reviewed by experts in the field, after which 101 papers were accepted for inclusion in these proceedings. Accepted papers are divided into nine groups, although there could be some overlaps. Articles written by the plenary, keynote, and invited speakers are also included in the proceedings.

We wish to express our appreciation to the Program Committee and additional reviewers who worked hard to ensure the quality of the contributions of this volume. We are thankful to Prof. M.A. Girolami for accepting our invitation to be the keynote speaker in this conference. We take this opportunity to express our gratitude to Profs. A. Skowron and S. Bramhachari for agreeing to be the plenary speakers of the conference. We thank Profs. D. Dasgupta, J. Udupa, N. Ganguly, P.K. Saha, S. Kuznetsov, and S.-B. Cho for accepting our invitation to deliver invited lectures during the conference. We gratefully acknowledge Alfred Hofmann, Executive Editor, Computer Science Editorial, Springer-Verlag, Heidelberg, Germany, for extending his co-operation in the publication of the PReMI 2013 proceedings. Finally, we would like to thank all the contributors for their enthusiastic response.

Thanks are due to Indranil Dutta for taking the major load of the secretarial jobs.

We wish PReMI 2013 conference academically very productive and enjoyable, and believe that you will find the proceedings to be a valuable source of reference for your ongoing and future research.

October 2013

Pradipta Maji
Ashish Ghosh
M.N. Murty
Kuntal Ghosh
Sankar K. Pal

Organization

PReMI 2013 is organized by the Machine Intelligence Unit, Indian Statistical Institute (ISI) in Kolkata during December 10–14, 2013.

PReMI 2013 Conference Committee

General Chair

S.K. Pal	ISI, Kolkata, India

Program Chairs

P. Maji	ISI, Kolkata, India
A. Ghosh	ISI, Kolkata, India
M.N. Murty	IISc, Bangalore, India

Organizing Chairs

K. Ghosh	ISI, Kolkata, India
Subhra S. Ray	ISI, Kolkata, India

Tutorial Chairs

S.N. Biswas	ISI, Kolkata, India
S. Mitra	ISI, Kolkata, India

Special Session Chairs

S. Bandyopadhyay	ISI, Kolkata, India
P.S. Sastry	IISc, Bangalore, Kolkata, India

Industrial Liaison

Santanu Chowdhury	IIT, Delhi, India

International Liaison

Simon C.K. Shiu	Hong Kong Polytechnic University, Hong Kong, SAR China

Web Chair

Indranil Dutta	ISI, Kolkata, India

Advisory Committee

Andrzej Skowron, Poland
Anil Jain, USA
B.L. Deekshatulu, India
B. Yegnanarayana, India
C.R. Rao, USA
D. Dutta Majumder, India
David W. Aha, USA
Gabriella Sanniti di Baja, Italy
Lotfi A. Zadeh, USA
Nasser M. Nasrabadi, USA
Rama Chellappa, USA
Ronald R. Yager, USA
Witold Pedrycz, Canada

Program Committee

Tinku Acharya	Videonetics Technology Private Limited, Bangalore, India
Arunabha Adhikari	West Bengal State University, Kolkata, India
R. Venkatesh Babu	Indian Institute of Science, Bangalore, India
Aditya Bagchi	Indian Statistical Institute, Kolkata, India
Sanghamitra Bandyopadhyay	Indian Statistical Institute, Kolkata, India
Minakshi Banerjee	RCC Institute of Information Technology, Kolkata, India
Mohua Banerjee	Indian Institute of Technology, Kanpur, India
Haider Banka	Indian School of Mines, Dhanbad, India
Anupam Basu	Indian Institute of Technology, Kharagpur, India
Laxmidhar Behera	Indian Institute of Technology, Kanpur, India
Dinabandhu Bhandari	Heritage Institute of Technology, Kolkata, India
Alok Bhattacharya	Jawaharlal Nehru University, Delhi, India
Bhargab Bhattacharya	Indian Statistical Institute, Kolkata
Mahua Bhattacharya	ABV Indian Institute of Information Technology and Management, Gwalior, India
Sourangshu Bhattacharya	Indian Institute of Technology, Kharagpur, India
Chiranjib Bhattacharyya	Indian Institute of Science, Bangalore, India
Dhruba Bhattacharyya	Tezpur University, India
Malay Bhattacharyya	Kalyani University, India
Partha Bhowmick	Indian Institute of Technology, Kharagpur, India
Kanad Kishore Biswas	Indian Institute of Technology, Delhi, India

Sambhu Nath Biswas	Indian Statistical Institute, Kolkata, India
Smarajit Bose	Indian Statistical Institute, Kolkata, India
Roberto Cesar	University of Sao Paulo, Brazil
Pinakpani Chakrabarti	Bose Institute, Kolkata, India
Basabi Chakraborty	Iwate Prefectural University, Japan
Chandan Chakraborty	Indian Institute of Technology, Kharagpur, India
Goutam Chakraborty	Iwate Prefectural University, Japan
Mihir Chakraborty	CSCR, Insian Statistical Institute, Kolkata, India
Snehashish Chakraverty	National Institute of Technology, Rourkela, India
Bhabatosh Chanda	Indian Statistical Institute, Kolkata, India
Amitava Chatterjee	Jadavpur University, Kolkata, India
Atal Chaudhuri	Jadavpur University, Kolkata, India
Debasis Chaudhuri	IAC, DEAL, Dehradun, India
Subhasis Chaudhuri	Indian Institute of Technology, Mumbai, India
Santanu Chaudhury	Indian Institute of Technology, Delhi, India
Sung-Bae Cho	Yonsei University, Korea
Ananda Shankar Chowdhury	Jadavpur University, Kolkata, India
Nirmalya Chowdhury	Jadavpur University, Kolkata, India
Amit Das	Bengal Engineering and Science University, Shibpur, India
Partha Pratim Das	India Institute of Technology, Kharagpur, India
Sukanta Das	Bengal Engineering and Science University, Shibpur, India
Sukhendu Das	Indian Institute of Technology, Chennai, India
Swagatam Das	Indian Statistical Institute, Kolkata, India
Dipankar Dasgupta	University of Memphis, USA
B.S. Daya Sagar	Indian Statistical Institute, Bangalore, India
Lipika De	TCS Innovation Labs, Delhi, India
Rajat De	Indian Statistical Institute, Kolkata, India
Alok Deb	India Institute of Technology, Kharagpur, India
Satchidananda Dehuri	Fakir Mohan University, Balasore, India
V. Susheela Devi	Indian Institute of Science, Bangalore, India
Andries Engelbrecht	University of Pretoria, South Africa
Paolo Gamba	University of Pavia, Italy
Niloy Ganguly	Indian Institute of Technology, Kharagpur, India
Utpal Garai	Indian Statistical Institute, Kolkata, India
Anil K. Ghosh	Indian Statistical Institute, Kolkata, India
Ashish Ghosh	Indian Statistical Institute, Kolkata, India
Indira Ghosh	Jawaharlal Nehru University, Delhi, India
Kuntal Ghosh	Indian Statistical Institute, Kolkata, India
Susmita Ghosh	Indian Statistical Institute, Kolkata, India

Phalguni Gupta	Indian Institute of Technology, Kanpur, India
D.S. Guru	University of Myshore, India
Anindya Halder	North-Eastern Hill University, Shillong, India
Larry Hall	University of South Florida, USA
Francisco Herrera	University of Granada, Spain
Asif Ikbal	Indian Institute of Technology, Patna, India
Syeda Darakhshan Jabeen	CSCR, Insian Statistical Institute, Kolkata, India
C.V. Jawahar	International Institute of Information Technology, Hyderabad, India
Jayadeva	Indian Institute of Technology, Delhi, India
B. Jayaram	Indian Institute of Technology, Delhi, India
Samarjit Kar	National Institute of Technology, Durgapur, India
Karmeshu	Jawaharlal Nehru University, Delhi, India
Nikola Kasabov	Auckland University of Technology, New Zealand
Mario Koeppen	Kyushu Institute of Technology, Japan
Amit Konar	Jadavpur University, Kolkata, India
Ravi Kothari	IBM Research, India
Rajeev Kumar	Indian Institute of Technology, Kharagpur, India
Malay Kundu	Indian Statistical Institute, Kolkata, India
Brian Lovell	University of Queensland, Australia
Subhamoy Maitra	Indian Statistical Institute, Kolkata, India
Santi Prasad Maity	Bengal Engineering and Science University, Shibpur, India
Banshidhar Majhi	National Institute of Technology, Rourkela, India
Pradipta Maji	Indian Statistical Institute, Kolkata, India
Kaushik Majumdar	Indian Statistical Institute, Bangalore, India
Deba Prasad Mandal	Indian Statistical Institute, Kolkata, India
Francesco Masulli	Università di Genova, Italy
Saroj Meher	Indian Statistical Institute, Bangalore, India
Mandar Mitra	Indian Statistical Institute, Kolkata, India
Pabitra Mitra	Indian Institute of Technology, Kharagpur, India
Suman Mitra	Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat, India
Sushmita Mitra	Indian Statistical Institute, Kolkata, India
Animesh Mukherjee	Institute for Scientific Interchange, Italy
Dipti Prasad Mukherjee	Indian Statistical Institute, Kolkata
Jayanta Mukherjee	Indian Institute of Technology, Kharagpur, India
C.A. Murthy	Indian Statistical Institute, Kolkata, India

M.N. Murty	Indian Institute of Science, Bangalore, India
Mike Nachtegael	Ghent University, Belgium
Sarif Naik	Philips Research Asia, Bangalore, India
Pradipta Nanda	Institute of Technical Education and Research, Bhubaneswar, India
G.C. Nandi	International Institute of Information Technology Allahabad, India
Sukumar Nandi	Indian Institute of Technology, Kharagpur India
Y. Narahari	Indian Institute of Science, Bangalore, India
Mita Nasipuri	Jadavpur University, Kolkata, India
Bhabesh Nath	Tezpur University, India
Tomoharu Nnkashima	Osaka Prefecture University, Japan
Sergei Obiedkov	School of Applied Mathematics and Information Science, Moscow, Russia
Hari Om	Indian School of Mines, Dhanbad, India
Sudarshan Padhy	Institute of Mathematics and Applications, Bhubaneswar, India
Malay Pakhira	Kalyani Government Engineering College, Kalyani, India
Amita Pal	Indian Statistical Institute, Kolkata, India
Asim K. Pal	Indian Institute of Management, Kolkata, India
Debnath Pal	Indian Institute of Science, Bangalore, India
Nikhil Pal	Indian Statistical Institute, Kolkata, India
Pinak Pani Pal	Indian Statistical Institute, Kolkata, India
Rajarshi Pal	Institute for Development and Research in Banking Technology, Hyderabad, India
Sukomal Pal	Indian School of Mines, Dhanbad, India
Tandra Pal	National Institute of Technology, Durgapur, India
Umapada Pal	Indian Statistical Institute, Kolkata, India
B.K. Panigrahi	Indian Institute of Technology, Delhi, India
Swapan Parui	Indian Statistical Institute, Kolkata, India
Gabriella Pasi	University of Milano Bicocca, Italy
Dipti Patra	National Institute of Technology, Rourkela, India
Swarnajyoti Patra	Tezpur University, India
Siarry Patrick	Université Paris-Est Créteil, Paris
Alfredo Petrosino	University of Naples Parthenope, Italy
Dilip Pratihar	Indian Institute of Technology, Kharagpur, India
Arun K. Pujari	University of Hyderabad, India
Subhra Sankar Ray	Indian Statistical Institute, Kolkata, India
Siddheswar Roy	Monash University, USA
Sumantra Dutta Roy	Indian Institute of Technology, Delhi, India

Pankaj Kumar Sa	National Institute of Technology, Rourkela, India
Punam Saha	University of Iowa, USA
Sanjoy Kumar Saha	Jadavpur University, India
Sriparna Saha	Indian Institute of Technology, Patna, India
Kamal Sarkar	Jadavpur University, India
Sudeshna Sarkar	Indian Institute of Technology, Kharagpur, India
Debashis Sen	National University of Singapore, Singapore
Rudy Setiono	National University of Singapore, Singapore
B. Uma Shankar	Indian Statistical Institute, Kolkata, India
Shirish K. Shevade	Indian Institute of Science, Bangalore, India
Simon C.K. Shiu	Hong Kong Polytechnic University, Hong Kong, SAR China
Jaya Sil	Bengal Engineering and Science University, Shibpur, India
Sitabhra Sinha	Institute of Mathematical Sciences, Chennai, India
Dominik Slezak	Infobright, Canada
Tanmoy Som	Indian Institute of Technology (BHU), India
Ponnuthurai Suganthan	National Technological University, Singapore
Shamik Sural	Indian Institute of Technology, Kharagpur, India
Brijesh Verma	CQ University, Australia
Deepu Vijayasena	National Institute of Technology, Karnataka, India
Mohammadi Zaki	Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat, India

Additional Reviewers

Bag, Soumen	Datta, Aloke	Maiti, Saptaditya
Banerjee, Abhirup	Dogra, Debi	Maitra, Subhamoy
Banerjee, Somnath	Ekbal, Asif	Mishra, Niladri Shekhar
Basu, Tanmay	Garain, Utpal	Mondal, Ajoy
Biswas, Sovan	Gorai, Apurba	Murthy, K. Ramachandra
Chanda, Sukalpa	Gupta, Lalit	Murthy, Ramachandra
Chattopadhyay, Swarup	Handl, Julia	Mutyam, Madhu
Chattopadhyay, Tanushyam	Hazra, Joydev	Nag, Abhijit
Chowdhury, Manish	Jith, Nirmal	Pal, Jayanta
Das Gupta, Rahul	Kolya, Anup	Palaiahnakote, Shivakumara
Das, Ranajit	Kumar-Pratihar, Dilip	Patra, Braja Gopal
Das, Sudeb	Kundu, Partha	Patra, Swarnajyoti
Dasgupta, Parthasarathi	Kundu, Suman	
	Lohar, Pintu	

Paul, Sushmita
Prasanna, S.R. Mahadeva
Ramakanth, Avinash
Reddy, Mopuri
Roy, Moumita
Roy, Rahul
Roy, Shaswati
Seth, Subhendu
Siarry, P
Subudhi, Badri Narayan
Sur, Arijit
Tomar, Namrata

Sponsoring Organizations from Academia

- Center for Soft Computing Research: A National Facility, ISI, Kolkata
- International Association for Pattern Recognition (IAPR)
- International Rough Set Society (IRSS)
- Web Intelligence Consortium (WIC)
- Indian National Academy of Engineering (INAE), Kolkata Chapter

Message from the General Chair

I am happy to see that PReMI 2013, the fifth version of the International Conference on Pattern Recognition and Machine Intelligence, is back in Calcutta where it originated in 2005. Its third and fourth editions were held in Delhi and Moscow in 2009 and 2011, respectively, in conjunction with another international conference on rough sets, fuzzy sets, and granular computing, namely, RSFDGrC. Like previous versions, PReMI 2013 has a wider scope with quality research papers, and focused broadly, using both classic and modern computing paradigms, on different tasks of pattern recognition, machine learning, data mining, granular computing, and related disciplines with various real-life problems such as in bioinformatics, web intelligence, big data analysis, biometrics, document processing, data security, video processing, computer vision, and social network mining. The conference includes various invited sessions by eminent scientists as well as an industrial session, among others. Special issues of several international journals will be published after the conference. All these make the event an ideal platform for both theoretical and applied researchers as well as practitioners for collaborative research.

I take this opportunity to thank all the participants, including the keynote, plenary, and invited speakers, reviewers, and the members of different committees for making the event a grand success. Thanks are also due to the sponsors for their support of this endeavor, and to Springer for publishing the proceedings in their prestigious LNCS series.

I believe the participants would have an enjoyable and productive stay in Kolkata.

October 2013 Sankar K. Pal

Invited Talks

Bayesian Methods to Detect Dye-Labelled DNA Oligonucleotides in Multiplexed Raman Spectra

Mark A. Girolami

Department of Statistical Science
University College London
1-19 Torrington Place London, WC1E 7HB
m.girolami@ucl.ac.uk

Abstract. Recent advances in the development of technology based on Raman scattering as a chemical analytical technique have made it possible to detect spectral mixtures of multiple DNA sequences quantitatively. However, to exploit these techniques fully, inferential methodologies are required which can deconvolute the observed mixture and infer the composition of distinct DNA sequences in the overall composite. Inferring the component spectra is posed as a model selection problem for a bilinear statistical model, and the Markov chain Monte Carlo inferential methodology required is developed. In particular a Gibbs sampler and reversible jump Markov chain Monte Carlo methods are presented along with techniques based on estimation of the marginal likelihood. The results reported are particularly encouraging, highlighting that, for multiplexed Raman spectra, inference of the composition of original sequences in the mixture is possible to acceptable levels of accuracy. This statistical me thodology makes the exploitation of multiplexed surface-enhanced resonance Raman scattering spectra in disease identification a reality.

Genetic Learning Algorithms in Developing a Framework for Cloud Security Insurance

Dipankar Dasgupta

Director, Center for Information Assurance
Professor of Computer Science
University of Memphis
Memphis, Tennessee 38152

Abstract. Cloud environments are becoming increasingly more attractive because of the possibilities in significant cost reduction in IT operations. Due to rapidly changing Information Technologies, it becomes more expensive to regularly update hardware and software licenses and also of having a big IT department with highly technical professional. It is being more challenging for government/industries to best use of their IT budget while securing the computing environment. Many organizations are adopting cloud services so that they can reduce the costs and increase the flexibility of their IT infrastructures. While different industry and sectors are moving to the cloud services for their IT need, they are also very concern about data security (both in storage and in transmission) and various compliance requirements such as PCI DSS, HIPAA, GLBA, SOX, ISO, etc. As cloud services move to the mainstream to meet major computing needs, the issues of ownership and chain of custody of customer data are becoming primary responsibilities of providers. The 3 fundamental cloud service models form a hierarchy, with Software-as-a-Service (SaaS) on top of Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) on the bottom. While security requirements are essential for all service models, they vary in degree of defensive measures. The compliance for each sector requires specific protection for online data such as Transparency, Respect for Context, Security, Focused Collection, Accountability, Access and Accuracy. For example, HIPAA (Health Insurance Portability and Accountability Act) requires insurance portability, administrative simplification and fraud enforcement like privacy and security. Another example, PCI-DSS compliance, was set up to improve the Information Security of financial transactions related to credit and debit cards. And GLBA (Gramm-Leach-Bliley Act) compliance requires analyzing the risks before moving customer information into emerging technology models.

This talk will cover various aspects of cloud computing (opportunities, issues, and challenges), the market view of cloud computing in government, public sectors, and security, as well as compliance issues and implications. We developed a software tool for Cloud Security (called MEGHNAD); if

time permits I will give a short demo of the tool. Here are two relevant research publications:

References

1. Dipankar Dasgupta, Moshiur Rahman. A Framework for Estimating Security Coverage for Cloud Service Insurance. In the proceedings of 7th Cyber Security and Information Intelligence Research Workshop. Oak Ridge, October 12–14, 2011.
2. Dipankar Dasgupta, Moshiur Rahman. Estimating Security Coverage for Cloud Services. In the proceedings of IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing,MIT. Boston: October 8-11, 2011.

Identifying Topics, Topical Experts and Users Interested in Specific Topics in Twitter

Niloy Ganguly

Department of CSE, IIT Kharagpur
niloy@cse.iitkgp.ac.in

The Twitter online social network (OSN) is now a popular platform for discovering real-time news on various topics. We are developing methodologies to improve topical search and recommendation in Twitter, which means searching for/recommending popular content or experts on specific topics of interest to the users. The primary challenge in these tasks is to accurately discover the topical attributes of individual users in Twitter, such as the topics of expertise or topics of interest of the users. We have proposed a novel methodology for this, which utilizes crowd-sourced social annotations of users, i.e., how the general Twitter population describes or annotates individual users. The social annotations are collected through the Twitter Lists feature. Utilizing the social annotations, we have developed a novel *who-is-who* system for Twitter [1], which gives the topical attributes of a specified user. We verified that the proposed methodology gives accurate and comprehensive topical attributes for millions of popular Twitter users.

The proposed methodology of discovering topical attributes of millions of Twitter users has numerous practical applications

1. We have developed a search system for topical experts in Twitter [2], utilizing the social annotations. Comparison of our system with the expert search service offered by Twitter shows that the proposed method provides better results for a large number of topical queries.
2. We have identified that the content posted by these experts is a valuable source of information on a large variety of topics. Specifically, all prior attempts to developing search/recommender systems on Twitter, considered random samples of the tweet stream. We have shown that compared to random samples, the content sampled from topical experts is far more content-rich and hence more suitable for such Information Retrieval and Data Mining tasks [3].
3. Most importantly, since our technique relies on crowd-sourcing, it enabled discovery of experts on thousands of topics in which Twitter users are interested. Whereas it is a general notion today that Twitter is associated with a few popular topics, such as music, politics, technology, and celebrity gossip, we, for the first time, showed that Twitter also consists of small groups of experts and users interested on a wide variety of niche, specialized topics ranging from neurology and anthropology to astrophysics and forensics [4].

Collaborators

- Krishna P. Gummadi, Max Planck Institute for Software Systems, Germany
- Naveen Kr. Sharma, IIT Kharagpur, India (presently at University of Washington, USA)
- Parantapa Bhattacharya, IIT Kharagpur, India
- Saptarshi Ghosh, IIT Kharagpur (presently at BESU Shibpur, India)
- Muhammad Bilal Zafar, Max Planck Institute for Software Systems, Germany
- Mainack Mondal, Max Planck Institute for Software Systems, Germany
- Juhi Kulshrestha, Max Planck Institute for Software Systems, Germany

References

1. Sharma, N., Ghosh, S., Benevenuto, F., Ganguly, N., Gummadi, K.: Inferring Who-is-Who in the Twitter Social Network. In: Proc. ACM Workshop on Online Social Networks (2012)
2. Ghosh, S., Sharma, N., Benevenuto, F., Ganguly, N., Gummadi, K.: Cognos: crowdsourcing search for topic experts in microblogs. In: Proc. ACM SIGIR (2012)
3. Ghosh, S., Zafar, M.B., Bhattacharya, P., Sharma, N., Ganguly, N., Gummadi, K.: On Sampling the Wisdom of Crowds: Random vs. Expert Sampling of the Twitter Stream In: Proc. ACM CIKM (2013)
4. Bhattacharya, P., Ghosh, S., Kulshrestha, J., Mondal, M., Zafar, M.B., Ganguly, N., Gummadi, K.: Deep Twitter Diving: Exploring Topical Groups in Microblogs at Scale. In: Proc. ACM CSCW. (2014)

Body-Wide Automatic Anatomy Recognition in Medical Imagery via Fuzzy Models

Jayaram K. Udupa

Professor of Radiologic Science in Radiology
Department of Radiology, University of Pennsylvania
Philadelphia, PA 19104-6021, USA
jay@mail.med.upenn.edu

Abstract. To make Quantitative Radiology (QR) a reality in radiological practice, computerized body-wide automatic anatomy recognition (AAR) becomes essential. With the goal of building a general AAR system that is not tied to any specific organ system, body region, or image modality, we present an AAR methodology for localizing and delineating all major organs in different body regions based on fuzzy modeling ideas and a tight integration of fuzzy models with an Iterative Relative Fuzzy Connectedness (IRFC) delineation algorithm. The methodology consists of five main steps: (a) gathering image data for building models and testing the AAR methodology from patient image sets existing in our health system; (b) formulating precise definitions of each body region and organ and delineating them following these definitions; (c) building hierarchical fuzzy anatomy models of organs for each body region; (d) recognizing and locating organs in given images by employing the hierarchical models; and (e) delineating the organs following the hierarchy. In Step (c), we explicitly encode object size and positional relationships into the hierarchy and subsequently exploit this information in object recognition in Step (d) and delineation in Step (e). Modality-independent and dependent aspects are carefully separated in model encoding. Furthermore, at the model building stage, a learning process is carried out for rehearsing an optimal threshold-based object recognition method. The recognition process in Step (d) starts from large, well-defined objects and proceeds down the hierarchy in a global to local manner. A fuzzy model-based version of the IRFC algorithm is created by naturally integrating the fuzzy model constraints into the delineation algorithm.

The AAR methodology is tested on three body regions thorax (on CT), abdomen (on CT and MRI), and neck (on MRI and CT) involving a total of over 35 organs and 130 data sets. The training and testing data sets are divided into equal size in all cases except for the neck. Overall the AAR method achieves a mean accuracy of about 2 voxels in localizing non-sparse blob-like objects and most sparse tubular objects. The delineation accuracy in terms of mean false positive and negative volume fractions is 2% and 8%, respectively, for non-sparse objects, and 5% and 15%, respectively, for sparse objects. The two object groups achieve mean boundary distance relative to ground truth of 0.9 and 1.5 vox-

els, respectively. We conclude that separation of modality-independent from dependent aspects, organization of objects in a hierarchy, encoding of object relationship information explicitly into the hierarchy, optimal threshold-based recognition learning, and fuzzy model-based IRFC are effective concepts which allowed us to demonstrate the feasibility of a general AAR system that works in different body regions on a variety of organs and on different modalities.

Table of Contents

Invited Papers

Pattern Recognition

Machine Learning

Image Processing

Speech and Video Processing

Medical Imaging

Document Image Processing

Soft Computing

Bioinformatics and Computational Biology

Social Media Mining

Interactive Computations: Toward Risk Management in Interactive Intelligent Systems

Andrzej Jankowski[1], Andrzej Skowron[2,⋆], and Roman Swiniarski[3]

[1] Institute of Computer Science, Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
a.jankowski@ii.pw.edu.pl
[2] Institute of Mathematics, University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
skowron@mimuw.edu.pl
[3] Department of Computer Science, San Diego State University
5500 Campanile Drive San Diego, CA 92182, USA
and
Institute of Computer Science Polish Academy of Sciences
Jana Kazimierza 5, 01-248 Warsaw, Poland
rswiniarski@mail.sdsu.edu

Abstract. Understanding the nature of interactions is regarded as one of the biggest challenges in projects related to complex adaptive systems. We discuss foundations for interactive computations in Interactive Intelligent Systems (IIS), developed in the Wistech program and used for behavior modeling of complex systems. We emphasize the key role of risk management in problem solving by IIS. The considerations are supported by real-life projects concerning, e.g., medical diagnosis and therapy support, control of an unmanned helicopter, algorithmic trading or fire commander decision support.

Keywords: rough sets, granular computing, interactive computations, interactive intelligent systems, risk management.

Traditional statistics is strong in devising ways of describing data and inferring distributional parameters from sample. Causal inference requires two additional ingredients: a science-friendly language for articulating causal knowledge, and a mathematical machinery for processing

⋆ This work was supported by the Polish National Science Centre grants 2011/01/B/ST6/03867, 2011/01/D/ST6/ 06981, and 2012/05/B/ST6/03215 as well as by the Polish National Centre for Research and Development (NCBiR) under the grant SYNAT No. SP/I/1/77065/10 in frame of the strategic scientific research and experimental development program: "Interdisciplinary System for Interactive Scientific and Scientific-Technical Information" and the grant No. O ROB/0010/ 03/001 in frame of the Defence and Security Programmes and Projects: "Modern engineering tools for decision support for commanders of the State Fire Service of Poland during Fire & Rescue operations in the buildings".

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 1–12, 2013.

that knowledge, combining it with data and drawing new causal conclusions about a phenomenon .

– *Judea Pearl* [15]

1 Introduction

Information granules (infogranules, for short) are widely discussed in the literature (see, *e.g.*, [16]). In particular, let us mention here the rough granular computing approach based on the rough set approach and its combination with other approaches to soft computing. However, the issues related to interactions of infogranules with the physical world and to perception of interactions in the physical world represented by infogranules are not well elaborated yet. On the other hand the understanding of interactions is the critical issue of complex systems [4] in which computations are progressing by interactions among information granules and physical objects.

We extend the existing approach to granular computing by introducing *complex granules* (c-*granules*, for short) [8] making it possible to model interactive computations performed by agents in Interactive Intelligent Systems (IIS) used for behavior modeling of complex systems.

Any agent operates in a local world of c-granules. The agent control is aiming to control computations performed on c-granules from this local world for achieving the target goals.

Computations in IIS are based on c-granules. The risk management in IIS is of the great importance for the success of behaviors of individuals, groups and societies of agents. The risk management tasks are considered as control tasks aiming at achieving the satisfactory performance of (societies of) agents. The novelty of the proposed approach is the use of complex vague concepts as the guards of control actions. These vague concepts are represented, *e.g.*, using domain ontologies. The rough set approach in combination with other soft computing approaches is used for approximation of the vague concepts relative to attributes (features) available to the risk management systems.

This paper is organized as follows. In Section 2 an introduction to Interactive Rough Granular Computing (IRGC) is presented. Issues related to reasoning based on adaptive judgement are included in Section 3. The approach to risk management based on IRGC is discussed in Section 4.

This paper covers some issues presented in the plenary talk at the 5th International Conference on Pattern Recognition and Machine Intelligence (PReMi 2013), December 10-14, 3013, Kolkata, India.

2 Interactive Rough Granular Computing (IRGC)

The essence of the proposed approach is the use of IIS implemented using IRGC [7,20,21,19,8,18]. The approach is based on foundations for modeling of IRGC relevant for IIS in which computations are progressing through interactions [4].

In IRGC interactive computations are performed on objects called *complex granules* (*c-granules*, for short) linking information granules [16] (or infogranules, for short) with physical objects called hunks [5,8].

Infogranules are widely discussed in the literature. They can be treated as specifications of compound objects (such as complex hierarchically defined attributes) together with scenarios of their implementations. Such granules are obtained as the result of information granulation [27]:

> *Information granulation can be viewed as a human way of achieving data compression and it plays a key role in implementation of the strategy of divide-and-conquer in human problem-solving.*

Infogranules belong to the concepts playing the main role in developing foundations for AI, data mining and text mining [16]. They grew up as some generalizations from fuzzy sets [25,26,27], rough set theory and interval analysis [16]. The rough set approach is crucial because of necessity to deal with approximations of inforgranules by the others, *e.g.*, in inducing classifiers for complex vague concepts. The IRGC is based on the rough set approach in combination with other approaches to soft computing (such as fuzzy sets). However, the issues related to interactions of infogranules with the physical world and their relation to perception of interactions in the physical world are not well elaborated yet [4,24]. On the other hand the understanding of interactions is the critical issue of complex systems [10]:

> [...] *interaction is a critical issue in the understanding of complex systems of any sorts: as such, it has emerged in several well-established scientific areas other than computer science, like biology, physics, social and organizational sciences.*

We propose to model complex systems by IIS created by societies of agents. Computations in the discussed IIS are based on c-granules [8] (see Figure 1). Any c-granule consists of three components, namely soft_suit, link_suit and hard_suit. These components are making it possible to deal with such abstract objects from soft_suit as infogranules as well as with physical objects from hard_suit. The link_suit of a given c-granule is used as a kind of c-granule interface for handling interaction between soft_suit and and hard_suit.

Calculi of c-granules are defined by elementary c-granules (determined, *e.g.*, by indiscernibility of similarity classes) and c-granules making it to possible to generate new c-granules from already defined ones (see Figure 1 where the presented c-granule produces new output c-granules from the given input c-granules). The hierarchy of c-granules is illustrated in Figure 2. Moreover, c-granules create the basis for the agent (communication) language construction and the language evolution.

Any agent operates in a local world of c-granules. The agent control is aiming to control computations performed on c-granules from this local world for achieving the target goals. Actions (sensors or plans) from link_suits of c-granules are used by the agent control in exploration and/or exploitation of the environment

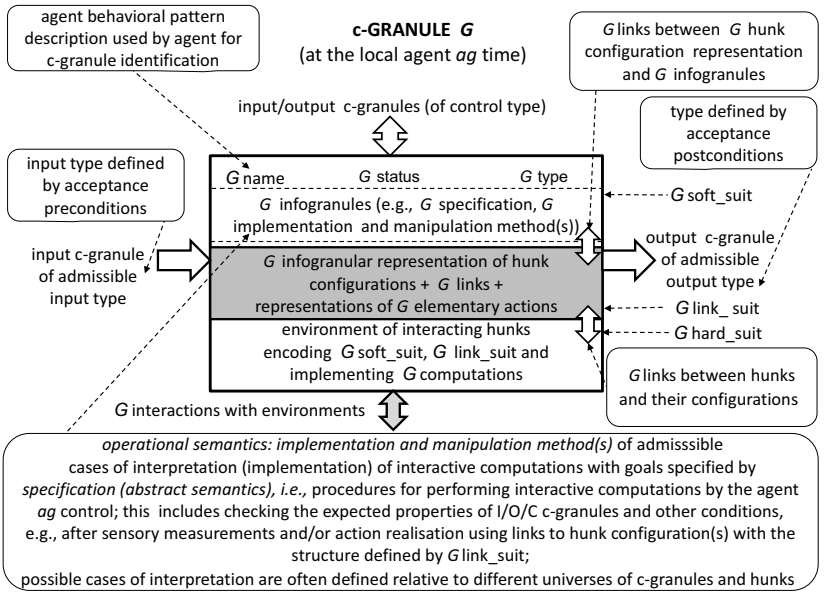


Fig. 1. General structure of c-granule

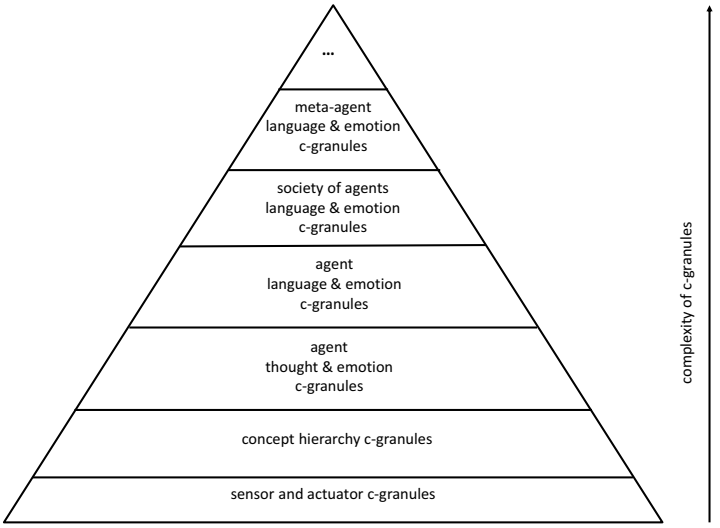


Fig. 2. Hierarchy of c-granules

on the way to achieve their targets. C-granules are also used for representation of perception by agents of interactions in the physical world. Due to the bounds of the agent perception abilities usually only a partial information about the

interactions from physical world may be available for agents. Hence, in particular the results of performed actions by agents can not be predicted with certainty. For more details on IRGC based on c-granules the reader is referred to [8].

One of the key issues of the approach to c-granules presented in [8] is some kind of integration of investigation of physical and mental phenomena. The integration follows from suggestions presented by many scientists. For illustration let us consider following two quotations strongly related to the research on IRGC based on c-granules:

> *As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.*
>
> – *Albert Einstein* ([3])

> *Constructing the physical part of the theory and unifying it with the mathematical part should be considered as one of the main goals of statistical learning theory.*
>
> – *Vladimir Vapnik* ([24], p. 721)

A special role in IRGC play information (decision) systems from the rough set approach [11,12,13,23]. They are used to record processes of interacting configurations of hunks. In order to represent interactive computations (used, *e.g.*, in searching for new features) information systems of a new type, namely interactive information systems, are needed [20,21,8].

3 Adaptive Judgement

The reasoning making it possible to derive relevant information granules for solutions of the target tasks is called *adaptive judgment*. *Intuitive judgment* and *rational judgment* are distinguished as different kinds of [9]. Among the tasks for adaptive judgment are the following ones supporting reasoning toward: searching for relevant approximation spaces, discovery of new features, selection of relevant features (attributes), rule induction, discovery of inclusion measures, strategies for conflict resolution, adaptation of measures based on the minimum description length principle, reasoning about changes, selection of parameters of (action and sensory) attributes, adaptation of quality measures over computations relative to agents, adaptation of object structures, discovery of relevant context, strategies for knowledge representation and interaction with knowledge bases, ontology acquisition and approximation, learning in dialogue of inclusion measures between information granules from different languages (*e.g.*, the formal language of the system and the user natural language), strategies for adaptation of existing models, strategies for development and evolution of communication language among agents in distributed environments, strategies for risk management in distributed computational systems.

Adaptive judgement in IIS is a mixture of reasoning based on deduction, abduction, induction, case based or analogy based reasoning, experience, observed

changes in the environment, meta-heuristics from natural computing (see Figure 3). We would like to stress that still much more work should be done

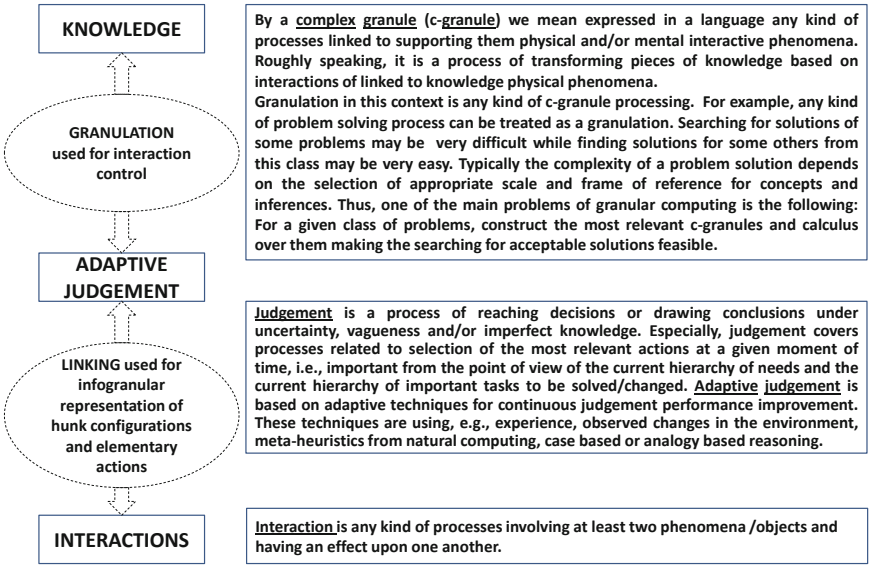


Fig. 3. Interactions, adaptive judgement and granulation

to develop approximate reasoning methods about complex vague concepts for making progress in development of IIS, in particular for the risk management in IIS. This idea was very well expressed by Leslie Valiant[1]:

> *A fundamental question for artificial intelligence is to characterize the computational building blocks that are necessary for cognition. A specific challenge is to build on the success of machine learning so as to cover broader issues in intelligence. [...] This requires, in particular a reconciliation between two contradictory characteristics – the apparent logical nature of reasoning and the statistical nature of learning.*

It is worthwhile to mention two more views. The first one by Lotfi A. Zadeh, the founder of fuzzy sets and the computing with words paradigm (see [26] and also `http://www.cs.berkeley.edu/~zadeh/presentations.html`):

> *Manipulation of perceptions plays a key role in human recognition, decision and execution processes. As a methodology, computing with words*

[1] see, *e.g.*, `http://en.wikipedia.org/wiki/Vagueness`, `http://people.seas.harvard.edu/~valiant/researchinterests.htm`

provides a foundation for a computational theory of perceptions - a theory which may have an important bearing on how humans make - and machines might make - perception-based rational decisions in an environment of imprecision, uncertainty and partial truth. [...] computing with words, or CW for short, is a methodology in which the objects of computation are words and propositions drawn from a natural language.

and the view by Judea Pearl included as the motto of this paper.

The question arises about the logic relevant for the above discussed tasks. First let us observe that the satisfiability relations in the IRGC framework can be treated as tools for constructing new information granules. If fact, for a given satisfiability relation, the semantics of formulas relative to this relation is defined. In this way the candidates for new relevant information granules are obtained. We would like to emphasize a very important feature. The relevant satisfiability relation for the considered problems is not given but it should be induced (discovered) on the basis of a partial information encoded in information (decision) systems. For real-life problems, it is often necessary to discover a hierarchy of satisfiability relations before we obtain the relevant target level. Information granules constructed at different levels of this hierarchy finally lead to relevant ones for approximation of complex vague concepts related to complex information granules expressed in natural language (see Figure 4).

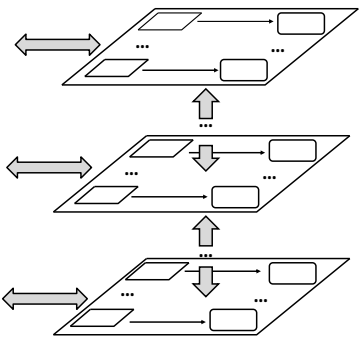


Fig. 4. Interactive hierarchical structures (gray arrows show interactions between hierarchical levels and the environment, arrows at hierarchical levels point from information (decision) systems representing partial specifications of satisfiability relations to induced from them theories consisting of rule sets)

4 Risk Management in IIS

Since the very beginning, all human activities were done at risk of failure. Recent years have shown the low quality of risk management in areas such as finance, economics, and many others. In this context, improvement in the risk

management has a particular importance for the further development of complex systems. The importance of risk management illustrates the following example from financial sector. Many of financial risk management experts consider Basel II rules[2] as a causal factor in the credit bubble prior to the 2007-8 collapse. Namely, in Basel II one of the principal factors of financial risk management was

> *outsourced to companies that were not subject to supervision, credit rating agencies.*

Of course, now we do have a new "improved" version of Basel II, called Basel III. However, according to an OECD [3] *the medium-term impact of Basel III implementation on GDP growth is negative and estimated in the range of* -0.05% *to* -0.15% *per year* (see also [22]).

On the basis of experience in many areas, we have now many valuable studies on different approaches to risk management. Currently, the dominant terminology is determined by the standards of ISO 31K [1]. However, the logic of inferences in risk management is dominated by the statistical paradigms, especially by Bayesian data analysis initiated about 300 years ago by Bayes, and regression data analysis initiated by about 200 years ago by Legendre and Gauss. On this basis, resulted many detailed methodologies specific for different fields. A classic example is the risk management methodology in the banking sector, based on the recommendations of Basel II standards for risk management mathematical models [17]. The current dominant statistical approach is not satisfactory because it does not give effective tools for inferences about the vague concepts and relations between them (see the included before sentences by L. Valiant).

A particularly important example of the risk management vague concept relation is the relation of a cause - effect relationships between various events. It should be noted that the concept of risk in ISO 31K is defined as *the effect of uncertainty on objectives*. Thus, by definition, the vagueness is also an essential part of the risk concept. To paraphrase the motto of this study by Judea Pearl, we can say that traditional statistical approach to risk management inference *is strong in devising ways of describing data and inferring distributional parameters from sample*. However, in practice risk management inference requires two additional ingredients (see the motto of this article):

- *a science-friendly language for articulating risk management knowledge*, and
- *a mathematical machinery for processing that knowledge, combining it with data and drawing new risk management conclusions about a phenomenon.*

Adding both mentioned above components is an extremely difficult task and binds to the core of AI research very accurately specified by the Turing test. With regard to our applications, properly adapted version of the test boils down to the fact that on the basis of a "conversation" with a hidden risk management expert and a hidden machine one will not be able to distinguish who is the man and who is the machine.

[2] see `http://en.wikipedia.org/wiki/Basel_Committee_on_Banking_Supervision`
[3] see `http://en.wikipedia.org/wiki/Basel_III`

We propose to extend the statistical paradigm by adding the two discussed components for designing of the high quality risk management systems supported by IIS.

For the risk management in IIS one of the most important task is to develop strategies for inducing approximations of vague complex concepts making it possible to check their satisfiability (to a degree). A typical example of such vague concept is the statement of the form: "now we do have very risky situation". The development of strategies for inducing approximations of such vague complex are based on the activation of actions performed by agents.

These vague complex concepts are represented by the agent hierarchy of needs. In risk management one should consider a variety of complex vague concepts and relations between them as well as reasoning schemes related, *e.g.*, to the bow-tie diagram (see Figure 5).

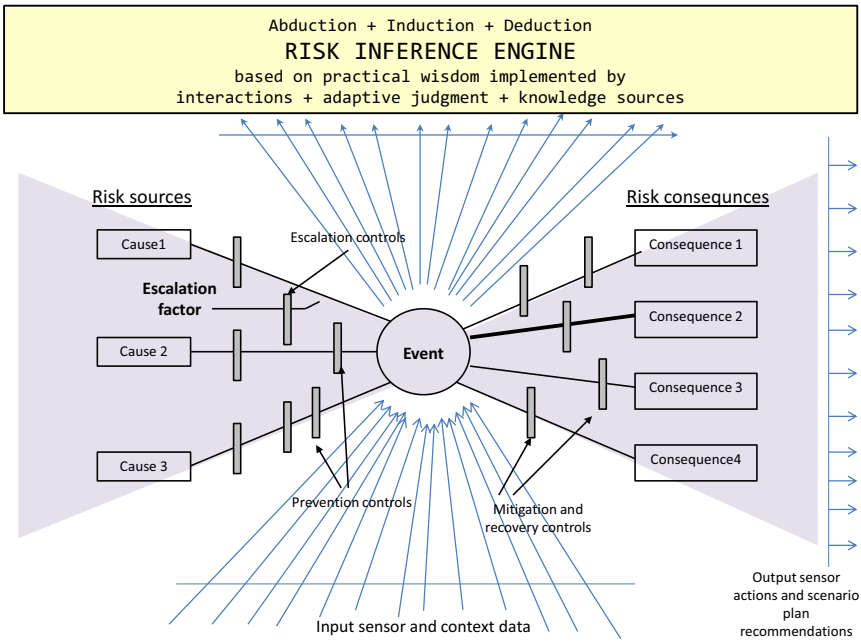


Fig. 5. Bow-tie diagram

Let us explain the bow-tie diagram using the chess game. Of course, the chess game is a very simple example. In practice the game could be much more complex. The bow-tie diagram has 3 basic parts:

1. concepts from risk sources,
2. current situation description represented by a hierarchy of concepts defined by the input sensors and context data,
3. concepts from risk consequences.

To make the next move in chess game the player should understand the current situation. To do this, he or she should use the domain knowledge representation (especially, related to the domain of risk management) and apply the relevant inference rules to the current situation description (see parts 1 and 2) enriched by knowledge about the history of moves. Based on the knowledge about possible sources of risk (expressed in part 1) and features of moves history, one should identify the prioritized list of hypotheses about the opposite player strategy. If the opposite player strategy is identified then it is much easier to win. This kind of inference leading to a list of the most likely to be true hypotheses for the opposite player strategy, is called abduction. This *is a form of logical inference that goes from observation to a hypothesis that accounts for the reliable data (observation) and seeks to explain relevant evidence* (by Wikipedia `http://en.wikipedia.org/wiki/Abductive_reasoning`). In the following step, the best possible next move should be proposed on the basis of the list of hypotheses for the opposite player strategy. For the chess game, one can generate the tree of all possible n-moves and propose the best next move using some well known algorithms (such as $minimax$, $alpha-beta$, $A-star$ [14]). In real life applications, such trees theoretically could be generated using the part of risk ontology related to consequences (part 3). If these trees are becoming huge then using relevant abduction inference one can try to identify constraints helping to make searching for the best next move in such trees feasible.

One can consider the mentioned above tasks of approximation of vague complex concepts initiating actions as the complex game discovery task (see Figure 6) from data and domain knowledge. The agents use the discovered games

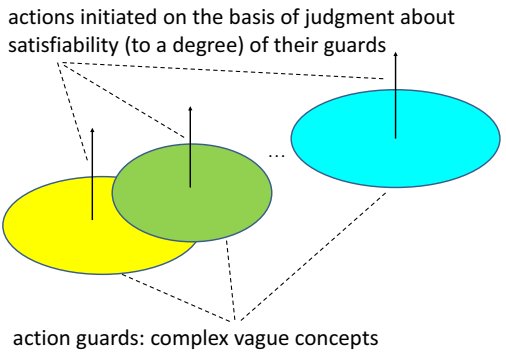


Fig. 6. Games based on complex vague concepts

for achieving their targets in the environment. The discovery process often is based on hierarchical learning supported by domain knowledge [8,2]. It is also worthwhile mentioning that such games are evolving in time (drifting in time) together with data and knowledge about the approximated concepts and the relevant strategies for adaptation of games used by agents are required. These

adaptive strategies are used to control the behavior of agents toward achieving by them targets. Note that also these strategies should be learned from available uncertain data and domain knowledge.

The discussed concepts such as interactive computation and adaptive judgment are among the basic ingredient elements in the Wisdom Technology (WisTech) [6,8]. Let us mention here the WisTech meta-equation:

$$\begin{aligned} \text{WISDOM} \;=\; & \\ & \text{INTERACTIONS} \;+ \\ & \text{ADAPTIVE JUDGEMENT} \;+ \\ & \text{KNOWLEDGE} \;. \end{aligned} \tag{1}$$

An extension of the rough set approach on interactive computations realized by IIS is one of the current challenges.

5 Conclusions

The approach for modeling interactive computations based on c-granules was presented and its importance for the risk managements was outlined.

The presented approach seems also to be of some importance for developing computing models in different areas such as natural computing (*e.g.*, computing models for meta-heuristics or computations models for complex processes in molecular biology), computing in distributed environments under uncertainty realized by multi-agent systems (*e.g.*, in social computing), modeling of computations for feature extraction (constructive induction) for approximation of complex vague concepts, hierarchical learning, discovery of planning strategies or strategies for coalition formation by IIS as well as for approximate reasoning about interactive computations based on such computing models.

References

1. ISO 31000 standard, `http://webstore.ansi.org/`
2. Bazan, J.: Hierarchical classifiers for complex spatio-temporal concepts. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 474–750. Springer, Heidelberg (2008)
3. Einstein, A.: Geometrie und Erfahrung (Geometry and Experience). Julius Springer, Berlin (1921)
4. Goldin, D., Smolka, S., Wegner, P. (eds.): Interactive Computation: The New Paradigm. Springer (2006)
5. Heller, M.: The Ontology of Physical Objects. Four Dimensional Hunks of Matter. Cambridge Studies in Philosophy. Cambridge University Press (1990)
6. Jankowski, A., Skowron, A.: A WisTech paradigm for intelligent systems. In: Peters, J.F., Skowron, A., Düntsch, I., Grzymała-Busse, J.W., Orłowska, E., Polkowski, L. (eds.) Transactions on Rough Sets VI. LNCS, vol. 4374, pp. 94–132. Springer, Heidelberg (2007)

7. Jankowski, A., Skowron, A.: Wisdom technology: A rough-granular approach. In: Marciniak, M., Mykowiecka, A. (eds.) Bolc Festschrift. LNCS, vol. 5070, pp. 3–41. Springer, Heidelberg (2009)
8. Jankowski, A., Skowron, A.: Practical Issues of Complex Systems Engineering: Wisdom Technology Approach. Springer, Heidelberg (2014) (in preparation)
9. Kahneman, D.: Maps of bounded rationality: Psychology for behavioral economics. The American Economic Review 93, 1449–1475 (2002)
10. Omicini, A., Ricci, A., Viroli, M.: The multidisciplinary patterns of interaction from sciences to computer science. In: Goldin, et al. (eds.) [14], pp. 395–414
11. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
12. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Knowledge Engineering and Problem Solving, vol. 9. Kluwer Academic Publishers, Dordrecht (1991)
13. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Information Sciences 177(1), 3–27 (2007)
14. Pearl, J.: Heuristics: Intelligent Search Strategies for Computer Problem Solving. The Addison Wesley, Moston (1984)
15. Pearl, J.: Causal inference in statistics: An overview. Statistics Surveys 3, 96–146 (2009)
16. Pedrycz, W., Skowron, S., Kreinovich, V. (eds.): Handbook of Granular Computing. John Wiley & Sons, Hoboken (2008)
17. Shevchenko, P. (ed.): Modelling Operational Risk Using Bayesian Inference. Springer (2011)
18. Skowron, A., Jankowski, A., Wasilewski, P.: Risk management and interactive computational systems. Journal of Advanced Mathematics and Applications 1, 61–73 (2012)
19. Skowron, A., Stepaniuk, J., Swiniarski, R.: Modeling rough granular computing based on approximation spaces. Information Sciences 184, 20–43 (2012)
20. Skowron, A., Wasilewski, P.: Information systems in modeling interactive computations on granules. Theoretical Computer Science 412(42), 5939–5959 (2011)
21. Skowron, A., Wasilewski, P.: Interactive information systems: Toward perception based computing. Theoretical Computer Science 454, 240–260 (2012)
22. Slovik, P., Cournède.: Macroeconomic Impact of Basel III, Working Papers, vol. 844. OECD Economics Publishing, OECD Economics Department (2011), `http://www.oecd.org/eco/Workingpapers`
23. Stepaniuk, J.: Rough-Granular Computing in Knowledge Discovery and Data Mining. SCI, vol. 152. Springer, Heidelberg (2008)
24. Vapnik, V.: Statistical Learning Theory. John Wiley & Sons, New York (1998)
25. Zadeh, L.A.: Fuzzy sets and information granularity. In: Advances in Fuzzy Set Theory and Applications, pp. 3–18. North-Holland, Amsterdam (1979)
26. Zadeh, L.A.: From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions. IEEE Transactions on Circuits and Systems 45, 105–119 (1999)
27. Zadeh, L.A.: A new direction in AI: Toward a computational theory of perceptions. AI Magazine 22(1), 73–84 (2001)

Fuzzy Digital Topology and Geometry and Their Applications to Medical Imaging

Punam K. Saha

Department of Electrical and Computer Engineering
Department of Radiology
University of Iowa, Iowa City, Iowa, USA

Abstract. The primary end-goal of most medical imaging research program is to collect information about function and physiology of internal human organs or tissues through a variety of *in vivo* or *ex vivo* imaging techniques. Often, medical imaging techniques suffer from limited spatial and temporal resolution, noise, background-inhomogeneity, and other artifacts leading to fuzzy representations of target objects in acquired images. Digital topology and geometry play important roles in medical image processing either by expanding the scope of target information or by providing a strong theoretical foundation to a process enhancing its stability, fidelity, and efficiency. The notions of digital topology and geometry are often intertwined in medical imaging applications and sometime it is difficult to draw a dividing line between them. This paper presents recent advancements and overviews of theory and computation of several fuzzy digital topologic and geometric approaches and describe their applications to medical imaging. More specifically, this paper discusses topics related to three-dimensional simple points, local topological parameters, fuzzy skeletonization, characterization of local structures, and their applications to research and clinical studies.

Keywords: Fuzzy subset, digital imaging, topology, geometry, distance transform, simple point, skeletonization, topological classification, medical imaging.

1 Introduction

The primary end-goal of most medical imaging research program is to collect information about function and physiology of internal human organs or tissues through a variety of in vivo or ex vivo imaging techniques [1]. Often, medical imaging techniques suffer from limited spatial and temporal resolution, noise, background-inhomogeneity, and other artifacts leading to fuzzy representations of target objects in acquired images. Digital topology and geometry [2-7] play important roles in medical image processing either by expanding the scope of target information or by providing a strong theoretical foundation to a process enhancing its stability, fidelity, and efficiency. The notions of digital topology and geometry are often intertwined in medical imaging applications and sometime it is difficult to draw a dividing line between them. The term "digital topology" loosely refers to the use of mathematical topological properties and features such as

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 13–29, 2013.

connectedness and boundary in computer vision, graphics and image processing. Digital topology has been studied since the late 1960's, when it was introduced by Rosenfeld, whose early articles [2, 3, 8-10] played a major role in establishing the field. It has provided the theoretical foundations for important image processing operations such as object counting, thinning, boundary detection and contour filling etc. During early 1970s, the subject of digital geometry was brought into light when the challenges of defining a digital straight line or straightness of a digital line were introduced. Subsequently, researcher investigated other areas of digital geometry, including curvature, corners, arc length, perimeter, surface area, convexity, convex hull, distance transform etc.

Digital topology and geometry have been applied in medical imaging to solve different important task. For example, component labeling and border tracking have been applied in medical image visualization, manipulation, and analysis [11]. Digital connectivity and minimum cost path has been widely used in medical image segmentation [12-15]. Digital topology has been applied to characterize local structure, e.g., classification of plates and rods in a trabecular bone imaging. Topology has also been used for correction of anatomic structures [16] segmented in acquired images in the presence of noise, partial voluming, and other artifacts. Skeletonization [17-19] has been used to solve several medical imaging purposes, including, representation of anatomic structure, path planning, feature extraction, disease identification etc. This paper presents recent advancements and overviews of theory and computation of several fuzzy digital topologic and geometric approaches and describe their applications to medical imaging. More specifically, this paper discusses topics related to three-dimensional simple points [6, 20, 21], local topological parameters [7], fuzzy skeletonization [19] , characterization of local structures [7, 22-25], and their applications to medical research and clinical studies.

2 Definitions and Notations

Most medical imaging techniques acquire scan data in a three-dimensional rectangular grid. A three dimensional (3-D) rectangular grid may be constructed by dividing the 3-D continuous space R^3 into rectangular parallelepipeds with three orthogonal families each of equally spaced parallel planes. The set of the centers of these rectangular parallelepipeds generates a rectangular grid and it is not difficult to see that, under a proper coordinate system, these points constitute the set Z^3 where Z is the set of all integers. Each upright rectangular parallelepiped centered at a point in Z^3 is referred to as a voxel or as a point. Two voxels are said to be 26-adjacent if the two share at least a vertex. Two voxels are said to be 18-adjacent if the two share at least an edge. Two voxels are said to be 6-adjacent if the two share a 2-D face. A κ-adjacency, where $\kappa \in \{6,18,26\}$, is defined in a way such that each point is κ-adjacent to exactly κ other points in the respective grid. A κ-path π is a nonempty sequence of points where every two successive points are κ-adjacent. Now consider

$S \subset Z^3$. Two points $x, y \in S$ is said to be κ-connected in S if there exists a κ-path from x to y that is contained by S. A κ-component of S is a maximal subset of S where every two points is κ-connected in S; S is said to be κ-connected if it has just one κ-component, or if S is an empty set. The set of 26-neighbors of a voxel p together with the voxel p forms a $3 \times 3 \times 3$ neighborhood and it is denoted by $N_{26}(p)$; $N_{26}^*(p)$ denotes the set of all voxels of $N_{26}(p)$ excluding the central voxel p. The other neighborhoods, namely, $N_6(p)$, $N_6^*(p)$, $N_{18}(p)$, and $N_{18}^*(p)$ are defined similarly.

Following Kong and Rosenfeld [4], a binary digital image or an object in a rectangular grind may be represented as a quadruple $\mathcal{B} = (Z^3, \alpha, \beta, B)$, where the first element Z^3 defines the 3-D image space; the two parameters α and β denote the adjacency relations used for object and background voxels, respective; here 26- and 6-adjacencies are used for object background voxels, respectively. Similarly, a fuzzy digital image or fuzzy object may be denoted as $\mathcal{O} = (Z^3, \alpha, \beta, \mu_{\mathcal{O}})$ where $\mu_{\mathcal{O}}: Z^3 \rightarrow [0,1]$ is the membership function of the target fuzzy object. The support O of the fuzzy object $\mathcal{O}$ is the set of all points with nonzero membership values, i.e., $O = \{p \mid p \in Z^3 \wedge \mu_{\mathcal{O}}(p) > 0\}$. In a fuzzy digital image, the adjacency relation α is used for voxels inside the support of the object while β is used for the voxels inside the background $\bar{O} = Z^3 - O$.

3 Fuzzy Distance Transform

Distance transform [26-28] is a local depth measure inside an object and we call it "fuzzy distance transform" when the object representation is fuzzy. Fuzzy distance transform [29] accounts for both partial voxel occupancy and spatial heterogeneity of an object distribution. The length of a link $\langle p, q \rangle$ is calculated as $\frac{1}{2}\left(\mu_{\mathcal{O}}(p) + \mu_{\mathcal{O}}(q)\right) \| p - q \|$, where, $\| \cdot \|$ denotes any Euclidean L2 norm. The length of a path $\pi = \langle p_1, p_2, \cdots, p_l \rangle$ in a fuzzy object $\mathcal{O}$, denoted by $\Pi_{\mathcal{O}}(\pi)$, is defined as the sum of lengths of all links along the path, i.e.,

$$\Pi_{\mathcal{O}}(\pi) = \sum_{i=1}^{l-1} \frac{1}{2}\left(\mu_{\mathcal{O}}(p_i) + \mu_{\mathcal{O}}(p_{i+1})\right) \| p_i - p_{i+1} \|. \tag{1}$$

Between any two points $p, q \in \mathbb{Z}^3$, infinitely many paths may exist; let $\mathcal{P}(p, q)$ denote the set of all paths from p to q. A path $\pi_{p,q} \in \mathcal{P}(p, q)$ is one of the shortest paths from p to q in $\mathcal{O}$, if and only if, $\Pi_{\mathcal{O}}(\pi_{p,q}) \leq \Pi_{\mathcal{O}}(\pi)$ for all $\pi \in \mathcal{P}(p, q)$. It may be noted that the shortest path between two points $p, q \in \mathbb{Z}^3$ may not be unique. The fuzzy distance from $p \in \mathbb{Z}^3$ to $q \in \mathbb{Z}^3$ in an object $\mathcal{O}$, denoted by $\omega_{\mathcal{O}}(p, q)$, is the length of one of the shortest paths from p to q, i.e.,

$$\omega_{\mathcal{O}}(p, q) = \min_{\pi \in \mathcal{P}(p,q)} \Pi_{\mathcal{O}}(\pi). \tag{2}$$

It has been shown in [29] that fuzzy distance satisfies the metric properties in both continuous as well as discrete spaces in any dimension. The fuzzy distance transform or FDT of an object $\mathcal{O}$ is represented as an image $\{(p, \Omega_{\mathcal{O}}(p)) | p \in \mathbb{Z}^3\}$, where $\Omega_{\mathcal{O}}: \mathbb{Z}^3 \rightarrow \Re^+$ denotes the fuzzy distance transform value at a given point; $\Re^+$ is the set of positive real numbers including zero. $\Omega_{\mathcal{O}}(p)$ is defined as the fuzzy distance between p and its nearest points in $\bar{\mathcal{O}}(\mathcal{O}) = \mathbb{Z}^3 - \mathcal{O}(\mathcal{O})$. In other words,:

$$\Omega_{\mathcal{O}}(p) = \min_{q \in \bar{\mathcal{O}}(\mathcal{O})} \omega_{\mathcal{O}}(p, q). \tag{3}$$

4 Simple Point and Topology Preservation

Saha et al. [6, 20, 21] conceived the concise characterization of 3-D simple points and presented it in the following four-condition format. A voxel or a point p is a (26,6) simple point if and only if it satisfies the following four conditions

Condition Saha 1: p has a white (background) 6-neighbor.

Condition Saha 2: p has a black (object) 26-neighbor.

Condition Saha 3: The set of black 26-neighbors of p is 26-connected.

Condition Saha 4: The set of white 6-neighbors of p is 6-connected in the set of its white 18-neighbors.

The above characterization of 3-D simple point was first documented in a technical report [20] published at the Indian Statistical Institute, February, 1991 which was communicated to Malandain [30] and a seminar talk [31] on the same topic was presented in June of 1991 at the INRIA, Sophia Antipolis Cedex, France. However, as claimed by Malandain [32], he failed to read the technical report and rediscovered the

same result which was presented as follows [33] – a voxel or a point p is a (26,6) 3-D simple point if and only if it satisfies the following two conditions.

Condition MB 1: $N^*(p)$ has exactly one 26-component of black points.

Condition MB 2: The number of 6-components of white points in $N_{18}^*(p)$ that intersect with $N_6^*(p)$ is exactly one.

It is straight forward that Condition MB 1 is equivalent to Conditions Saha 1 and 3 while Condition MB 2 is equivalent to Conditions Saha 2 and 4 which was later acknowledged by Bertrand and Malandain [34]. In fact, the four conditions used in Saha *et al.*'s 3-D simple point characterization define the four different types of topological violations which occur when a non-simple voxel/point is deleted. Condition Saha 1 ensures that no cavity is created by deletion of a point while Condition Saha 2 confirms that no isolated point is deleted. Condition Saha 3 is used to ensure that, after the deletion of a point p, its neighboring points remains connected. Finally, Condition Saha 4 guarantees that the deletion of a point does not create a tunnel in its neighborhood. The following two theorems by Saha *et al.* [6, 20, 21] are crucial to solve the fundamental challenges of 3-D simple point characterization.

Theorem 1. If point p has at a white 6-neighbor, the number of tunnels $\eta(p)$ in $N_{26}^*(p)$ is one less than the number of 6-components of white points in $N_{18}^*(p)$ that intersect with $N_6^*(p)$, or, zero otherwise.

Theorem 2. The number of cavities in $N_{26}^*(p)$ is one when all 6-neighbors of p is black (object) and zero otherwise.

The above two theorems define the following local topological numbers proposed by Saha *et al.* [6, 7, 20, 21] that relate to different topological entities in the local neighborhood of a voxel/point characterizing its topological properties.

$\xi(p)$: The number of objects components in $N_{26}^*(p)$

$\eta(p)$: The number of tunnels in $N_{26}^*(p)$

$\delta(p)$: The number of cavities $N_{26}^*(p)$

Betrand and Malandain [34, 35] have advocated the following topological numbers.

$T_{26}(p)$: The number of 26-components of black points in $N_{26}^*(p)$

$T_6(p)$: The number of 6-components of white points in $N_{18}^*(p)$ that intersect with $N_6^*(p)$

Although, the number $T_6(p)$ is inspired by Theorems 1 and 2, it fails to distinguish between the two different topological situations – creation of a tunnel versus creation of a cavity.

5 Fuzzy Skeletonization

Blum's pioneering work on grassfire transform [36] led to the notion of skeletonization that converts a volumetric object into a union of surfaces and curves. The process is defined using fire propagation on a grass field, where the field resembles an object. The fire is simultaneously set at all boundary points and it propagates inwardly at a uniform speed. The skeleton is defined as the set of quench points where two or more fire fronts meet. However, the notion of skeletonization for fuzzy objects has not yet been defined. To define a fuzzy skeletonization process, we suggest modifying the Blum's grassfire transform for a fuzzy object where the membership function is interpreted as local material density so that the speed of grassfire at a given point is inversely proportional to its material density. Following this notion, it can be shown that fuzzy distance transform (FDT) [29] value at a point p is proportional to the time when the fire front reaches p. Therefore, during the fuzzy grassfire propagation, the speed of a fire front at a point equates to the inverse of local material density and this equality is violated only at quench points where the propagation process is interrupted. Thus, a voxel $p \in Z^3$, where Z is the set of integers and Z^3 represents a rectangular image grid, is a fuzzy quench voxel in a fuzzy digital object $\mathcal{O}$ if the following inequality holds for every neighbor q of p

$$FDT(q) - FDT(p) < \frac{1}{2}\left(\mu_{\mathcal{O}}(p) + \mu_{\mathcal{O}}(q)\right)|p - q|. \tag{4}$$

Saha and Wehrli [37] introduced the above definition of fuzzy quench voxel which was further studied by Svensson [38] where she referred to it as the center of fuzzy maximal ball (CFMB). Also, it may be noted that the definition of fuzzy quench voxel is equivalent to that of center of maximal ball (CMB) [39] for binary digital objects. Two types of quench points may form –surface- and curve-quench points (Figure 1). A surface quench point is formed when two opposite fire fronts meet while a curve quench point is formed when fire fronts meet from all directions on a plane. In a digital space, surface-quench voxel is formed when two opposite fire fronts meet along x-, y- or z-direction and a curve-quench voxel is formed when fire fronts meet from all eight directions in xy-, yz-, or zx-planes.

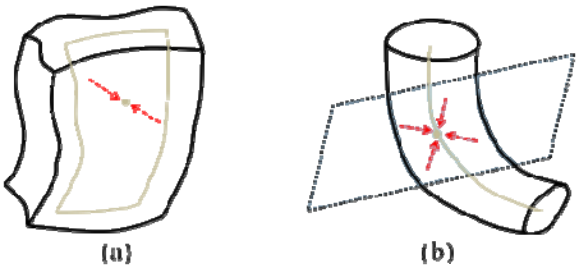


Fig. 1. Illustration of independent fire fronts meeting at surface- (a) and curve-like (b) quench voxels

During the process of fuzzy skeletonization, voxels are removed in the increasing order of their FDT values. The overall fuzzy skeletonization process is summarized in the following. Here, $O = \{p \mid \mu_O(p) > 0\}$ denotes the support of the fuzzy object O.

Primary skeletonization
select voxels $p \in O$ in the order of FDT values
if p is not a fuzzy axial voxel and its deletion preserves 3-D topology and 2D topology on mid-planes
remove p from O, i.e., set $\mu_O(p) = 0$

Final skeletonization:
select voxels $p \in O$ in the order of FDT values
if p is in two-voxel thick structure and its deletion preserves 3-D topology and 2D topology on mid-planes
remove p from O, i.e., set $\mu_O(p) = 0$
select voxels $p \in O$ in the order of FDT values
if topologic and geometric features of p fail to agree and its deletion preserves 3-D topology
remove p from O, i.e., set $\mu_O(p) = 0$

This notion of primary and final skeletonization was simultaneously introduced by Saha et al. [18] and Arcelli et al. [39]. Saha et al. described it in 3-D while Arcelli et al. presented the idea in 2-D for a DT-based skeletonization algorithm.

Removal of a voxel $p \in O$ preserves the topology of O if and only if p is a (26,6) simple voxel [6] in O. Beside the 3-D topology preservation condition, an additional constrain of 2D topology preservation in all three middle planes of the candidate voxel is subjected to ensure continuity of surface-like structures and to avoid undesired drilling effects as illustrated by Saha et al. [18].

Several new concepts including fuzzy axial voxels, local and global significance factors and two-voxel thick structures are introduced for fuzzy objects are presented by Jin and Saha [19]. Also, a condition defining the disagreement between topological and geometric features of a voxel is provided. Finally, a new noisy skeletal branches pruning algorithm based on global significance factor (GSF) is introduced [19].

Results of application of skeletonization and pruning of two other medical images are presented in Figure 2. For all these examples, the results of skeletonization and pruning are visually encouraging.

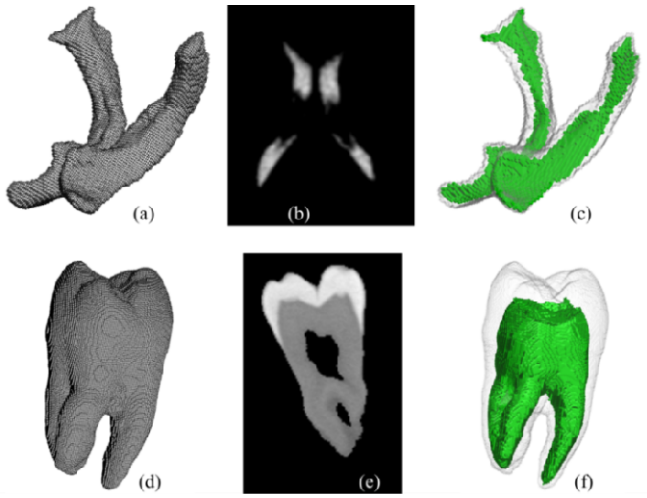


Fig. 2. Results of application of fuzzy skeletonization on two anatomic structures fuzzy segmented from acquired images. (a) A part of cerebrospinal fluid segmented from human brain MR imaging. (b) An axial image slice illustrating the fuzziness and noise. (c) Surface rendering of the fuzzy skeleton. (d-f) Same as (a-c) but for micro-CT data of a human tooth.

6 Characterization of Local Structures

In this section, two different approaches of characterizing local structures are presented. Digital topological analysis uniquely determine the topological class of each voxel in a skeletal representation of a 3-D object. On the other hand, in a volumetric representation of an object, the process volumetric topological analysis characterizes individual location on the continuum between a perfect plate and a perfect rod. These two different methods are described in the following.

7 Digital Topological Analysis

Digital topological analysis or DTA was invented by Saha and Chaudhuri [7] and later adopted in the context of trabecular bone image analysis [22, 23, 25]. For the sake of brevity, here, we only describe some basic ideas necessary to understand the developments in the following sections. Let us explain the idea on an object $\mathcal{S}$ that may be represented as a union of finitely many surfaces and curves in a continuous 3-D-space (see Figure 3(a)). It is interesting to note that a point in $\mathcal{S}$ may be classified depending on its local topological property. For example, the removal of a point p_s on

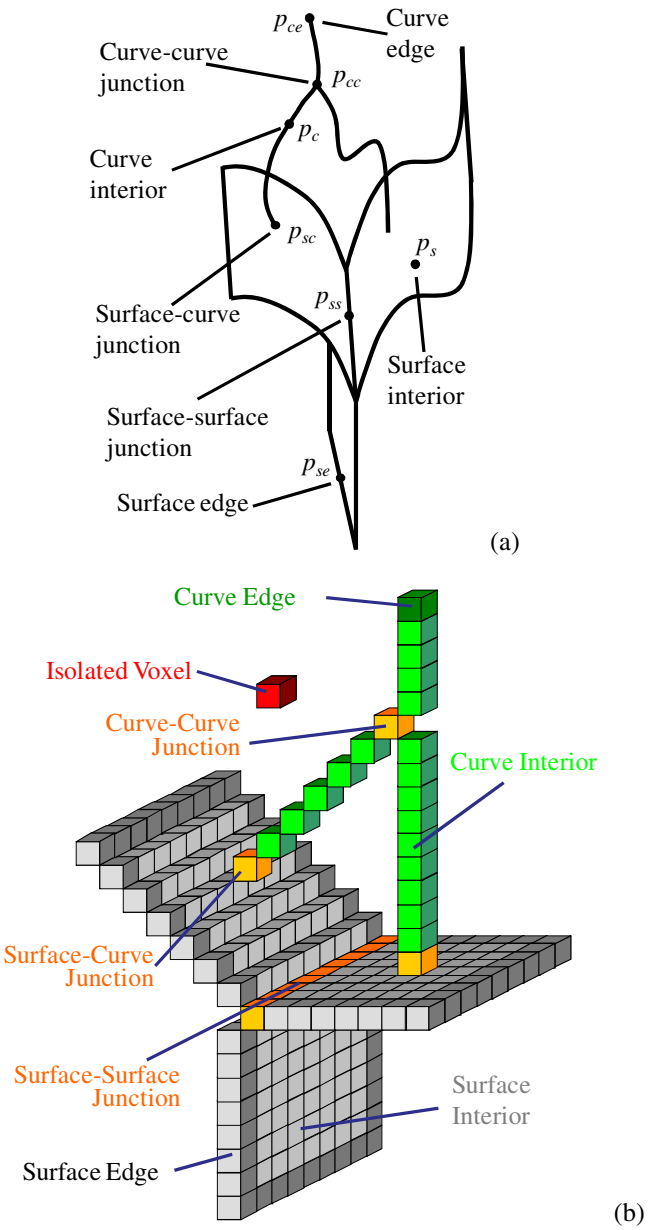


Fig. 3. Illustrations of possible topological classifications. (a) An example in the continuous 3-D space. (b) An example in a digital space.

a surface of $\mathcal{S}$ creates exactly one tunnel in a sufficiently small neighborhood of that point. The removal of a point p_{ss} on a junction of multiple surfaces creates multiple tunnels. It creates exactly two objects on a curve (p_c) and multiple objects at a junction of curves (p_{cc}). At a junction of surface and curve (p_{sc}), it creates multiple objects and exactly one tunnel and so on.

In a digital space, we identify eight different topological classes of skeletal voxels, namely, isolated, curve interiors, curve edges, surface interiors, surface edges, and curve-curve, surface-curve, and surface-surface junctions (Figure 3(b)). However, the classification in a digital space is not as straightforward as in the case of R^3. DTA is solved in three sequential steps – (1) determination of local topological type, (2) initial topological classification based on these types and (3) final classification after corrections at different topological junctions. Local topological type [7] provides only a partial classification which is essentially derived from the three local topological parameters $\xi(p)$, $\eta(p)$ and $\delta(p)$ (see Section 4). During the second step, a unique initial classification is achieved by reviewing the local topological type of 26-neighbors of each voxel not uniquely classified during the first step. After initial classification, the process is complete except for some possible corrections at junctions.

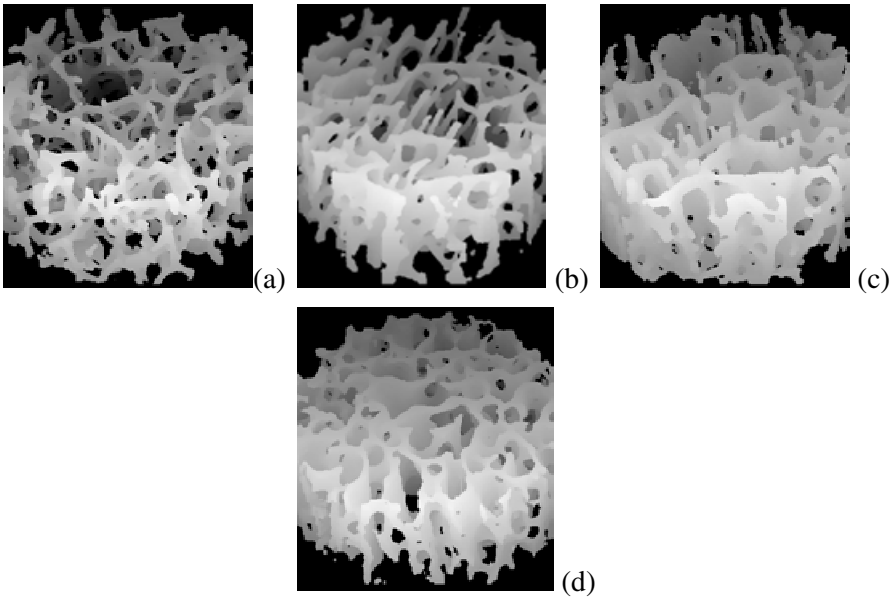


Fig. 4. Rod-like versus plate-like networks in 3-D projection μMRI images of radial TB [22]: (a) highly rod-like (F 74 yrs.), (b) moderately rod-like (F 53 years), (c) predominantly plate-like (M 68 yrs.), (d) highly plate-like (M 60 years). Topological parameters parallel visual assessment of increasing plate-like morphology with the surface-to-curve ratio increasing nearly 20-fold from (a) through (d).

This step is primarily accomplished by analyzing the topology of edge type voxels and the fact that, under a normal circumstance, edge voxels on a surface form a 26-closed curve and this normal topology of edge voxels is perturbed only at junctions. Local topological properties have been used by others researchers to decompose an object into different segments [35, 40].

This method has been widely applied in characterizing trabecular bone micro-architecture at in vivo resolution. Two composite parameters, the surface-to-curve ratio (S/C) and erosion index (EI) were derived from individual topological parameters obtained using digital topological analysis (DTA). EI, for example, was defined as the ratio of the sum of parameters expected to increase upon osteoclastic resorption, divided by the sum of parameters expected to decrease secondary to such processes. Figure 4 illustrates the sensitivity of the topological parameters to bone-loss induced structural changes: the BV/TV value of the four subjects varies by a factor of 2 while the S/C varies by a factor of 20 [22]. In an in vivo MR study [25] involving 79 post menopausal women covering a range of DEXA BMD's, topological entities (surface density, S/C) measured at the distal radius were found to be statistically significant discriminators of patients with vertebral deformities (n=29) from those without deformities. Over the last few years the DTA technique has been applied in several clinical and research studies at the University of Pennsylvania [25, 41] and has also been adopted by several other research groups [42-44].

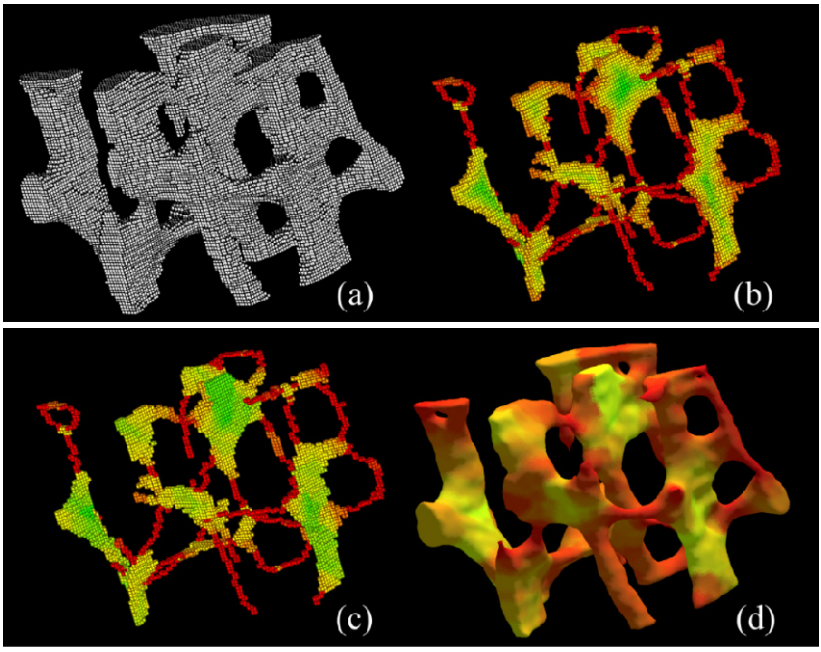


Fig. 5. Results of intermediate steps during the VTA process. (a) A TB region selected from a μCT image of a human ankle specimen. (b-d) Color coded (red: pure rod, green: o\pure plate, yellow: intermediate) voxel display of manifold distance transform (b), manifold scale (c) and surface rendition of VTA (d).

8 Volumetric Topological Analysis

Volumetric topological analysis (VTA) [24] identifies topological entities on the continuum between a perfect plate and a perfect rod in a volumetric BMD representation of a TB network. VTA starts with a fuzzy object O (e.g., TB) and is completed in five sequential steps (Figure 5) – (1) fuzzy surface skeletonization [19], (2) digital topological analysis (DTA) [7, 23], (3) manifold distance transform (MDT) [24], (4) manifold scale computation [24], and (5) volumetric feature propagation [24]. The output of VTA is a function $VTA: O \rightarrow R$ where $VTA(p)$ at an object voxel p gives the local object structure width measurement (μm); see Figure 5(d) for interpretation of different measurements on a surface-like volumetric structure. Two trabecular bone measures, namely, mean trabecular bone plate surface-width (SW_{VTA}) and trabecular surface to curve ratio (SCR_{VTA}) are computed using VTA analyses over a target volume of interest. Recently, tensor scale [45, 46] has been used to compute local structure width. Results from a cadaveric validation study on MDCT based trabecular and cortical bone measures demonstrated high repeat scan reproducibility of VTA measures. (Table 1(a)) and strong ability to predict experimental bone strength (Table 1(b), Figure 6).

Table 1. (a) Intraclass correlation (ICC) values of MDCT bone measures from three repeat cadaveric scans (N = 15). (b) Ability of MDCT TB measures to predict experimental bone strength; r values of linear correlation with TB yield stress. Abbreviations – TB: trabecular bone; Th: thickness; Poro: porosity; vBMD: volumetric BMD; SW: surface-width; NA: Network area; Sp: spacing

TB vBMD	TB SW	TB NA	TB Th	TB Sp
0.996	0.98	0.98	0.95	0.96

(a)

TB vBMD	TB SW	TB NA	TB Th	TB Sp
0.89	0.96	0.94	0.93	0.93

(b)

Results of Human Studies: An MDCT human bone imaging protocol (IRB ID # 201106785) was established at the University of Iowa. The use of ultra-sharp reconstruction kernels overcame the limitation of spatial resolution observed by previous investigators [47]. In a human pilot study [48], ten athletes (6 male), eleven patients (5 male) with cystic fibrosis (CF), 12 patients (6 male) on continuous treatment with a selective serotonin reuptake inhibitor (SSRI) for at least one year (duration: 3.3±3.0 years; fluoxetine-equivalent dose: 36±13 mg/day), and three patients (all female) with clinically confirmed anorexia nervosa (AN) were recruited (18 to 23 years of age for all groups). Age-sex-BMI-similar healthy controls were selected from the Iowa Bone Development Study (IBDS) cohort (42 male/46 female). MDCT scans of the distal

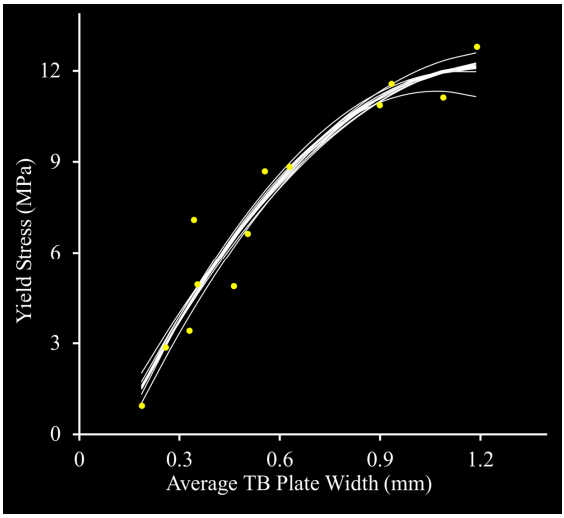


Fig. 6. A leave-one-out experiment demonstrated that MDCT derived TB plate width predicts experimental bone strength with an average error of 14%. The prediction functions for different instances of leave-one-out tests were stable.

tibia and whole body (WB) DXA were acquired. Pairwise comparisons between IBDS controls and participants from the three patient groups revealed that differences in cortical bone in SSRI and AN patients are non-significant, while CF patients demonstrate detectably lower cortical bone (Figure 7). The SSRI, CF and AN groups had compromised TB vBMD and TB microarchitecture which more prominently distinguished them from controls as compared to DXA BMD (Figure 7). The CF and AN groups had significantly higher TB spacing as compared to the SSRI group (Figure 7) indicating possible heterogeneous bone loss (Figure 8(c)) with potentially enhanced

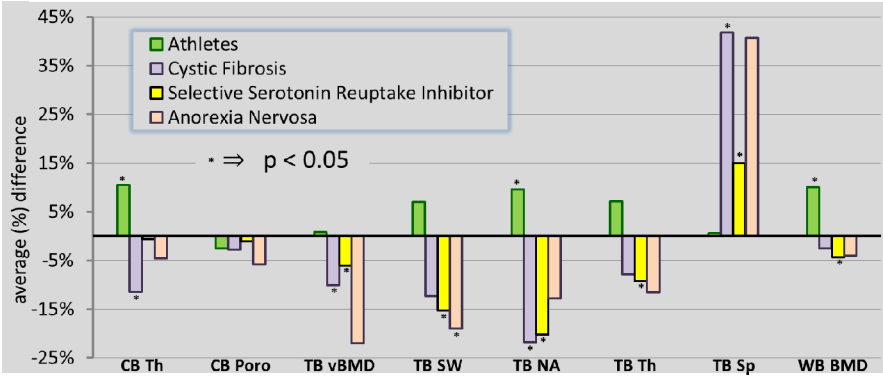


Fig. 7. Average differences of bone measures in athlete, cystic fibrosis, selective serotonin reuptake inhibitor, and anorexia nervosa groups as compared to age-sex-BMI-similar healthy controls from the Iowa Bone Development Study. Age-sex-height matching was used for the anorexia nervosa group.

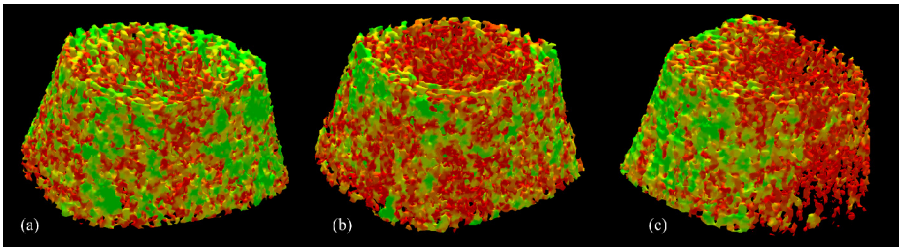


Fig. 8. Color-coded illustration of TB plate/rod classification for a IBDS female control (a) and an age-similar, sex- and BMI-matched patient on continuous treatment with an SSRI (b), and another age-similar, sex- and BMI-matched patient with confirmed diagnosis of CF (c). The healthy female (a) has more TB plates (green) as compared to the two patient participants. Between the two patients, the CF patient (c) has some signs of heterogeneous bone loss.

risks of fracture. In the athlete group, differences in cortical bone, compared to controls, predominated over differences in TB measures.

Peripheral MDCT imaging, together with advanced processing algorithms, are well-suited for in vivo skeletal assessment, capturing differences in volumetric, geometric, and microarchitectural properties of cortical and trabecular bones. Performance of the method needs to be validated against state-of-art HR-pQCT and the preliminary findings from the pilot in vivo study require confirmation in a larger sample accounting for various potential confounders.

9 Conclusion

Digital topology and geometry are useful in many real-life medical imaging applications benefitting clinical and research studies. Primarily, digital topology and geometry play three important roles at the level of method design and developments – (1) effective solution of many classical problems in medical imaging, (2) expansion of the scope of target information, and (3) provision of a strong theoretical foundation to a process enhancing its stability, fidelity, and efficiency. Over the last several decades there have been significant advancements in different areas of digital topology and geometry, including, distance analyses, topology preservation, skeletonization, local structure characterization etc. and some of these methods are being routinely applied on human studies. There are several unanswered fundamental questions related to digital topology and geometry which provides more research opportunities in this field.

Acknowledgements. This work was supported by the NIH grants R01-AR054439, R01-DE012101, and UL1-RR024979.

References

1. Cho, Z.H., Jones, J.P., Singh, M.: Foundation of Medical Imaging. Wiley, New York (1993)
2. Rosenfeld, A.: Fuzzy digital topology. Information and Control 40, 76–87 (1979)

3. Rosenfeld, A.: Thee dimensional digital topology. Information and Control 50, 119–127 (1981)
4. Kong, T.Y., Rosenfeld, A.: Digital topology: introduction and survey. Computer Vision, Graphics, and Image Processing 48, 357–393 (1989)
5. Klette, R., Rosenfeld, A.: Digital Geometry: Geometric Methods for Digital Picture Analysis. Morgan Kaufmann, San Francisco (2004)
6. Saha, P.K., Chaudhuri, B.B.: Detection of 3-D simple points for topology preserving transformations with application to thinning. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 1028–1032 (1994)
7. Saha, P.K., Chaudhuri, B.B.: 3D digital topology under binary transformation with applications. Computer Vision and Image Understanding 63, 418–429 (1996)
8. Rosenfeld, A.: Connectivity in digital pictures. Journal of the Association for Computing Machinery 17, 146–160 (1970)
9. Rosenfeld, A.: Adjacency in digital pictures. Information and Control 26, 24–33 (1974)
10. Rosenfeld, A.: Arcs and curves in digital Pictures. Journal of the Association for Computing Machinery 20, 81–87 (1973)
11. Udupa, J.K., Odhner, D.: Fast visualization, manipulation, and analysis of binary volumetric objects. IEEE Transactions on Computer Graphics and Applications 11, 53–62 (1991)
12. Falcão, A.X., Udupa, J.K., Samarasekera, S., Sharma, S.: User-steered image segmentation paradigms: live wire and live lane. Graphical Models and Image Processing 60, 233–260 (1998)
13. Barrett, W.A., Mortensen, E.N.: Interactive live-wire boundary extraction. Medical Image Analysis 1, 331–341 (1997)
14. Udupa, J.K., Samarasekera, S.: Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. Graphical Models and Image Processing 58, 246–261 (1996)
15. Saha, P.K., Udupa, J.K., Odhner, D.: Scale-based fuzzy connected image segmentation: theory, algorithms, and validation. Computer Vision and Image Understanding 77, 145–174 (2000)
16. Han, X., Xu, C., Braga-Neto, U., Prince, J.L.: Topology correction in brain cortex segmentation using a multiscale, graph-based algorithm. IEEE Trans. Med. Imaging 21, 109–121 (2002)
17. Arcelli, C., Sanniti di Baja, G., Serino, L.: Distance-driven skeletonization in voxel images. IEEE Trans. Pattern Anal. Mach. Intell. 33, 709–720 (2011)
18. Saha, P.K., Chaudhuri, B.B., Dutta Majumder, D.: A new shape preserving parallel thinning algorithm for 3D digital images. Pattern Recognition 30, 1939–1955 (1997)
19. Jin, D., Saha, P.K.: A new fuzzy skeletonization algorithm and its applications to medical imaging. In: Petrosino, A. (ed.) ICIAP 2013, Part I. LNCS, vol. 8156, pp. 662–671. Springer, Heidelberg (2013)
20. Saha, P.K., Chanda, B., Majumder, D.D.: Principles and algorithms for 2-D and 3-D shrinking. Indian Statistical Institute, Calcutta (1991)
21. Saha, P.K., Chaudhuri, B.B., Chanda, B., Dutta Majumder, D.: Topology preservation in 3D digital space. Pattern Recognition 27, 295–300 (1994)
22. Gomberg, B.G., Saha, P.K., Song, H.K., Hwang, S.N., Wehrli, F.W.: Topological analysis of trabecular bone MR images. IEEE Transactions on Medical Imaging 19, 166–174 (2000)
23. Saha, P.K., Gomberg, B.R., Wehrli, F.W.: Three-dimensional digital topological characterization of cancellous bone architecture. International Journal of Imaging Systems and Technology 11, 81–90 (2000)

24. Saha, P.K., Xu, Y., Duan, H., Heiner, A., Liang, G.: Volumetric topological analysis: a novel approach for trabecular bone classification on the continuum between plates and rods. IEEE Trans. Med. Imaging 29, 1821–1838 (2010)
25. Wehrli, F.W., Gomberg, B.R., Saha, P.K., Song, H.K., Hwang, S.N., Snyder, P.J.: Digital topological analysis of in vivo magnetic resonance microimages of trabecular bone reveals structural implications of osteoporosis. Journal of Bone Mineral Research 16, 1520–1531 (2001)
26. Borgefors, G.: Distance transform in arbitrary dimensions. Computer Vision Graphics Image Processing 27, 321–345 (1984)
27. Borgefors, G.: Distance transformations in digital images. Computer Vision Graphics and Image Processing 34, 344–371 (1986)
28. Borgefors, G.: On digital distance transformation in three dimensions. Computer Vision Graphics Image Processing 64, 368–376 (1996)
29. Saha, P.K., Wehrli, F.W., Gomberg, B.R.: Fuzzy distance transform: theory, algorithms, and applications. Computer Vision and Image Understanding 86, 171–190 (2002)
30. Malandain, G.: Acknowledgement of the techinical report by Saha et al titled - principles and algorithms for 2D and 3D shrinking. In: Sophia Antipolis Cedex, France, July 2 (1991); recipient: P. K. Saha
31. Saha, P.K.: 2D thinning algorithms and 3D shrinking. In: INRIA, Sophia Antipolis Cedex, France (June 1991)
32. Malandain, G.: Explantion of failure to cite the techinical report by Saha et al titled - principles and algorithms for 2D and 3D shrinking. In: Sophia Antipolis Cedex, France, April 7 (1993)
33. Malandain, G., Bertrand, G.: Fast characterization of 3-D simple points. In: Proceedings of 11th International Conference on Pattern Recognition, pp. 232–235 (1992)
34. Bertrand, G., Malandain, G.: A new characterization of three-dimensional simple points. Pattern Recognition Letters 15, 169–175 (1994)
35. Malandain, G., Bertrand, G., Ayache, N.: Topological segmentation of discrete surfaces. International Journal of Computer Vision 10, 183–197 (1993)
36. Blum, H. (ed.): A transformation for extracting new descriptors of shape (Models for the Perception of Speech and Visual Form). MIT Press, Cambridge (1967)
37. Saha, P.K., Wehrli, F.W.: Fuzzy distance transform in general digital grids and its applications. Presented at the 7th Joint Conference on Information Sciences, Research Triangular Park, NC (2003)
38. Svensson, S.: Aspects on the reverse fuzzy distance transform. Pattern Recognition Letters 29, 888–896 (2008)
39. Arcelli, C., Sanniti di Baja, G.: Well-shaped, stable, and reversible skeletons from the (3,4)-distance transform. Journal of Visual Communication and Image Representation 5, 107–115 (1994)
40. Serino, L., Sanniti di Baja, G., Arcelli, C.: Using the skeleton for 3D object decomposition. Presented at the Scandinavian Conference on Image Analysis, Ystad Saltsjöbad, Sweden (2011)
41. Ladinsky, G.A., Vasilic, B., Popescu, A.M., Wald, M., Zemel, B.S., Snyder, P.J., Loh, L., Song, H.K., Saha, P.K., Wright, A.C., Wehrli, F.W.: Trabecular structure quantified with the MRI-based virtual bone biopsy in postmenopausal women contributes to vertebral deformity burden independent of areal vertebral BMD. J. Bone Miner. Res. 23, 64–74 (2008)
42. Stauber, M., Muller, R.: Volumetric spatial decomposition of trabecular bone into rods and plates–a new method for local bone morphometry. Bone 38, 475–484 (2006)

43. Liu, X.S., Sajda, P., Saha, P.K., Wehrli, F.W., Guo, X.E.: Quantification of the roles of trabecular microarchitecture and trabecular type in determining the elastic modulus of human trabecular bone. J. Bone Miner. Res. 21, 1608–1617 (2006)
44. Chang, G., Pakin, S.K., Schweitzer, M.E., Saha, P.K., Regatte, R.R.: Adaptations in trabecular bone microarchitecture in Olympic athletes determined by 7T MRI. J. Magn. Reson. Imaging 27, 1089–1095 (2008)
45. Liu, Y., Saha, P.K., Xu, Z.: Quantitative characterization of trabecular bone microarchitecture using tensor scale and multi-detector CT imaging. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 124–131. Springer, Heidelberg (2012)
46. Saha, P.K.: Tensor scale: a local morphometric parameter with applications to computer vision and image processing. Computer Vision and Image Understanding 99, 384–413 (2005)
47. Burghardt, A.J., Link, T.M., Majumdar, S.: High-resolution computed tomography for clinical imaging of bone microarchitecture. Clin. Orthop. Relat. Res. 469, 2179–2193 (2011)
48. Saha, P.K., Amelon, R.E., Liu, Y., Li, C., Jin, D., Chen, C., Fishbaugher, J.M., Letuchy, E.M., Calarge, C.A., Janz, K.F., Hornick, D.B., Eichenberger-Gilmore, J., Burns, T.L., Torner, J.C., Levy, S.M.: In vivo study of trabecular and cortical bone in young adults with varying trajectories of bone development using multi-row detector CT imaging. In: Proceedings of Annual Meeting of the American Society for Bone and Mineral Research, Baltimore, MD (2013)

Scalable Knowledge Discovery in Complex Data with Pattern Structures

Sergei O. Kuznetsov

School of Applied Mathematics and Information Science,
National Research University Higher School of Economics,
Bol. Trekhsvyatitelskii 3, Moscow, Russia
skuznetsov@hse.ru

Abstract. Pattern structures propose a direct way to knowledge discovery in data with structure, such as logical formulas, graphs, strings, tuples of numerical intervals, etc., by defining closed descriptions and discovery tools build upon them: automatic construction of taxonomies, association rules and classifiers. A combination of lazy evaluation with projections of initial data, randomization and parallelization suggest efficient approach which is scalable to big data.

1 Introduction

In many real-world knowledge discovery problems researchers have to deal with complex descriptions different from binary datatables. In the last two decades the use of closed descriptions defined either in terms of Galois connections, semilattical similarity operation (i.e., operation which is idempotent, commutative, and associative) or in equivalent terms of counting inference proved to be very useful in various knowledge discovery applications, such as ontology and taxonomy engineering, mining association rules, machine learning, classification, and clustering. Several attempts were done in defining closed sets of graphs and closed graphs [24,32,26,30,28,1,18], strings [11,6], numerical intervals [23,22], logical formulas [7,10], etc. In [15] a general approach called pattern structures was proposed, which allows one to apply knowledge discovery tools to arbitrary partially ordered data descriptions. Using pattern structures, one can compute taxonomies, ontologies, implicational dependencies and their bases, association rules, and classifiers in the same way as it is done with binary data.

To meet the big data challenge the problem settings of knowledge discovery can be recast to allow for faster procedures. In this paper we show how the classification problems for pattern structures can be reformulated to achieve scalability even for complex descriptions.

The rest of the paper is organized as follows: In Section 2 we recall basic definitions in pattern structures, give examples of graph-based and interval-based pattern structures. In Section 3 we describe our approach to scalable classification with pattern structures, and make a conclusion in Section 4.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 30–39, 2013.

2 Knowledge Discovery with Pattern Structures

2.1 Main Definitions and Results

Let G be a set (of objects), let $(D, \sqcap)$ be a meet-semi-lattice (of all possible object descriptions) and let $\delta : G \to D$ be a mapping. Then $(G, \underline{D}, \delta)$, where $\underline{D} = (D, \sqcap)$, is called a *pattern structure*, provided that the set $\delta(G) := \{\delta(g) | g \in G\}$ generates a complete subsemilattice $(D_\delta, \sqcap)$ of $(D, \sqcap)$, i.e., every subset X of $\delta(G)$ has an infimum $\sqcap X$ in $(D, \sqcap)$. Elements of D are called *patterns* and are naturally ordered by subsumption relation $\sqsubseteq$: given $c, d \in D$ one has $c \sqsubseteq d \Leftrightarrow c \sqcap d = c$. Operation $\sqcap$ is also called a *similarity operation*. A pattern structure $(G, \underline{D}, \delta)$ gives rise to the following derivation operators $(\cdot)^\diamond$:

$$A^\diamond = \sqcap_{g \in A} \delta(g) \qquad \text{for } A \subseteq G,$$
$$d^\diamond = \{g \in G \mid d \sqsubseteq \delta(g)\} \qquad \text{for } d \in (D, \sqcap)$$

These operators form a Galois connection between the powerset of G and $(D, \sqsubseteq)$. The pairs (A, d) satisfying $A \subseteq G$, $d \in D$, $A^\diamond = d$, and $A = d^\diamond$ are called the *pattern concepts* of $(G, \underline{D}, \delta)$, with *pattern extent* A and *pattern intent* d. Pattern concepts are ordered wrt. set inclusion on extents. The ordered set of pattern concepts makes a lattice, called *pattern concept lattice*. For $a, b \in D$ the *pattern implication* $a \to b$ holds if $a^\diamond \subseteq b^\diamond$, and the *pattern association rule* $a \to_{c,s} b$ with *confidence* c and *support* s holds if $s \leq \frac{|a^\diamond \cap b^\diamond|}{|G|}$ and $c \leq \frac{|a^\diamond \cap b^\diamond|}{|a^\diamond|}$. Like in case of association rules [33,34], pattern association rules may be inferred from a concise representation that corresponds to the set of edges of the diagram of the pattern concept lattice. Operator $(\cdot)^{\diamond\diamond}$ is an algebraical closure operator on patterns, since it is idempotent, extensive, and monotone.

The concept-based learning model for standard object-attribute representation (i.e., formal contexts) [12,25,27] is naturally extended to pattern structures. Suppose we have a set of positive examples G_+ and a set of negative examples G_- w.r.t. a *target attribute*, $G_+ \cap G_- = \emptyset$, objects from $G_\tau = G \setminus (G_+ \cup G_-)$ are called undetermined examples.

A pattern $c \in D$ is a *positive premise (classifier)* iff

$$c^\diamond \cap E_- = \emptyset \text{ and } \exists A \subseteq E_+ : c \sqsubseteq A^\diamond$$

A pattern $h \in D$ is a *positive hypothesis* iff

$$h^\diamond \cap E_- = \emptyset \text{ and } \exists A \subseteq E_+ : h = A^\diamond$$

A positive hypothesis is the *least general generalization* of descriptions ("similarity") of positive examples, which is not contained in (does not cover) any negative example. *Negative* premises (classifiers) and hypotheses are defined similarly. Various classification schemes using premises are possible, as an example consider the following simplest scheme from [12,26,15]: If description $\delta(g)$ of an undetermined example g contains a positive premise (hypothesis) c, i.e., $c \sqsubseteq \delta(g)$,

and no negative premise, then g is *classified positively*. Negative classifications are defined similarly. If $\delta(g)$ contains premises (hypotheses) of both signs, or if $\delta(g)$ contains no premise (hypothesis) at all, then the classification is contradictory or undetermined, respectively, and some probabilistic relaxation of the above definitions of premises and hypotheses should be applied.

For some pattern structures (e.g., for the pattern structures on sets of graphs with labeled vertices) even computing subsumption of patterns may be NP-hard. Hence, for practical situations one needs approximation tools, which would replace the patterns with simpler ones, even if that results in some loss of information. To this end we use a contractive monotone and idempotent mapping $\psi : D \to D$ that replaces each pattern $d \in D$ by $\psi(d)$ such that the pattern structure $(G,\ \underline{D},\ \delta)$ is replaced by $(G,\ \underline{D},\ \psi \circ \delta)$. Under some natural algebraic requirements that hold for all natural projections in particular pattern structures we studied in applications, see [30], the meet operation $\sqcap$ is preserved: $\psi(X \sqcap Y) = \psi(X) \sqcap \psi(Y)$. This property of a projection allows one to relate premises (hypotheses) in the original representation with those approximated by a projection.The representation context of the projected case is obtained from the unprojected one by removing some attributes. If $\psi(a) \to \psi(b)$ and $\psi(b) = b$ then $a \to b$ for arbitrary $a,\ b \in D$. In particular, if $\psi(a)$ is a positive (negative) premise in projected representation, then a is positive (negative) premise in the original representation.

2.2 Pattern Structures in Applications

One may argue that a semi-lattice on descriptions is a too demanding requirement, but we can show easily that this is not the case. Any natural kind of descriptions available for data analysis has an explicitly or implicitly given partial order relation in the form of "is a" or "part of" relation. Having a partially ordered set $(P,\ \leq)$ of descriptions one can define a *similarity operation* $\sqcap$ on sets of descriptions as follows: For two descriptions X and Y from P

$$\{X\} \sqcap \{Y\}\ := \{Z | Z \leq X, Y,\ \forall Z_* \leq X, Y\ Z_* \not\geq Z\},$$

i.e., $\{X\} \sqcap \{Y\}$ is the set of all maximal common subdescriptions of descriptions X and Y. Similarity of non-singleton sets of descriptions $\{X_1,\ \ldots,\ X_k\}$ and $\{Y_1,\ \ldots,\ Y_m\}$ is defined as

$$\{X_1,\ \ldots,\ X_k\} \sqcap \{Y_1,\ \ldots,\ Y_m\} := MAX_{\leq}(\bigcup_{i,j}(\{X_i\} \sqcap \{Y_j\})),$$

where $MAX_{\leq}(\mathcal{X})$ returns maximal elements of $\mathcal{X}$ w.r.t. $\leq$. The similarity operation $\sqcap$ on sets of descriptions is commutative: $\mathcal{X} \sqcap \mathcal{Y} = \mathcal{Y} \sqcap \mathcal{X}$ and associative: $(\mathcal{X} \sqcap \mathcal{Y}) \sqcap \mathcal{Z} = \mathcal{X} \sqcap (\mathcal{Y} \sqcap \mathcal{Z})$. A set $\mathcal{X}$ of descriptions from P for which $\mathcal{X} \sqcap \mathcal{X} = \mathcal{X}$ defines a pattern. Then the triple $(G, (D,\ \sqcap), \delta)$, where D is the set of all patterns, is a pattern structure.

One can think of $\mathcal{X} \sqcap \mathcal{Y}$ in the following way, which also gives a straightforward approach to computing $\sqcap$: One takes the set of all subdesriptions of all

descriptions of $\mathcal{X}$ and takes set-theoretic intersection (i.e., $\cap$) of this set with the set of all subdescriptions of all descriptions of $\mathcal{Y}$. Finally, from the resulting set of subdescriptions one chooses the maximal ones w.r.t. the partial order $\leq$ on descriptions.

Pattern Structures on Sets of Graphs. In [24,26] we proposed a semi-lattice on sets of graphs with labeled vertices and edges. This semilattice is based on a partial order given by subgraph isomorphism or its generalizations. For example, in [26,15] the following natural order relation on graphs with labeled vertices and edges, called *domination relation*, was proposed. Consider connected graphs[1] with vertex and edge labels from set $\mathcal{L}$ partially ordered by $\preceq$. Denote the set of graphs with labeled vertices and edges by P. Each graph Γ from P is a quadruple of the form $((V,\ l),\ (E,\ b))$, where V is a set of vertices, E is a set of edges, $l : V \to \mathcal{L}$ is a function assigning labels to vertices, and $b : E \to \mathcal{L}$ is a function assigning labels to edges. In $(P,\ \leq)$ we do not distinguish isomorphic graphs.

For two graphs $\Gamma_1 := ((V_1,\ l_1),\ (E_1,\ b_1))$ and $\Gamma_2 := ((V_2,\ l_2),\ (E_2,\ b_2))$ from P we say that Γ_1 *dominates* Γ_2 or $\Gamma_2 \leq \Gamma_1$ (or Γ_2 is a *subgraph* of Γ_1) if there exists an injection $\varphi : V_2 \to V_1$ such that it *respects edges*: $(v, w) \in E_2 \Rightarrow (\varphi(v), \varphi(w)) \in E_1$ and *fits under labels*: $l_2(v) \preceq l_1(\varphi(v))$, if $(v, w) \in E_2$, then $b_2(v, w) \preceq b_1(\varphi(v), \varphi(w))$.

Obviously, $(P,\ \leq)$ is a partially ordered set. Having a partial order on graphs, one can use the definitions above to define similarity operation $\sqcap$ and closure operator $(\cdot)^{\diamond\diamond}$. A set of graphs X is called *closed* if $X^{\diamond\diamond} = X$. The closed set of graphs consists of *closed graphs* as defined in [36]. A learning model based on *graph pattern structures* along the lines of the previous subsection was used in series of applications in chemo- and bioinformatics [16,30], in text analysis [14] and conflict analysis [13]. Numerous types of projections were used in these applications, like e.g. k-vertex subgraphs, k-length paths, cylcic subgraphs in chemoinformatics, and noun, verb, and other types of phrases in natural language processing.

Pattern Structures on Intervals. In practice, a typical object-attribute data table is not binary, but has many-valued attributes. Instead of binarizing (scaling) data, one can directly work with many-valued attributes by applying *interval pattern structures*. For two intervals $[a_1,\ b_1]$ and $[a_2,\ b_2]$, with $a_1,\ b_1,\ a_2,\ b_2 \in \mathbb{R}$, we define their meet as

$$[a_1,\ b_1] \sqcap [a_2,\ b_2] = [\min(a_1,\ a_2),\ \max(b_1,\ b_2)].$$

This operator is obviously idempotent, commutative and associative, thus defining a pattern structure on tuples (vectors) of intervals of attribute values. The lattice of interval pattern structures is isomorphic to the concept lattice of the context that arises from the *interordinal scaling* of the initial many-valued numerical context, where for each table value a two binary attributes $\geq a$ and $\leq a$

[1] Omitting the condition of connectedness, one obtains a similar, but computationally much harder model.

are introduced. However, interval tuples give better understanding of results and computation with them is faster than that with the interordinal scaling, as shown in experiments with gene expression data [22].

3 Scalable Classification with Pattern Structures

The goal of computing implications, association rules, premises, hypotheses, and there concise representations is to "understand" data by creating "knowledge" in the form of implicational dependencies, and to use these dependencies for making predictions for new data. Intractability results on the sizes of concepts [27], implication bases [29,9,3], (minimal) hypotheses [27,2] say that the amount of "knowledge" generated from data without due filtering can be larger than data themselves. If one uses knowledge for making predictions, i.e., defining missing information, say classes of objects described by new data, one does not need having all "knowledge" given explicitly, one just needs having predictions equivalent to those made when all knowledge is there. To attain this goal, one can note that classification by means of premises can be done in a "lazy" way without explicit computation of all premises. Classification can be described in the extended pattern structure

$$(G, (D*, \sqcap*), \delta*) = (G, ((D, \sqcap) \times (\{0, 1\}, \wedge)), \delta \cup val),$$

where $\wedge$ is logical conjunction and the mapping $val : G \to \{0, 1\}$ says whether an object has the target attribute or not. Below we show how it works for premises in the classification scheme defined above.

Many algorithms for computing concept lattices, like NextClosure [17] and CbO [25], may be adapted to computing pattern lattices in bottom-up way. The worst-case time complexity of computing all pattern concepts of a pattern structure $(G,\ \underline{D},\ \delta)$ in the bottom-up way is $O((p(\sqcap) + p(\sqsubseteq)|G|) \cdot |G| \cdot |L|)$, where $p(\sqcap)$ is time needed to perform $\sqcap$ operation, $p(\sqsubseteq)$ is time needed to test $\sqsubseteq$ relation, and L is the set of all pattern concepts. In case of graphs, even $p(\sqsubseteq)$ may be exponential w.r.t. the number of graph vertices, that is why approximations (like those given by projections) are often needed. For a fixed projection size $p(\sqsubseteq)$ and $p(\sqcap)$ can be considered constant. To compute graph patterns in the top-bottom way, e.g., for computing *frequent patterns*, one can update CbO algorithm by getting access to the "fine" structure of descriptions, like it was done for graphs in [28]. The worst-case time complexity of computing the set of interval pattern structures is $O(|G|^2 \cdot |M| \cdot |L|)$, where M is the set of attributes, which in practice can be much lower than the worst-case complexity of computing the set L of all concepts of the interordinally scaled numerical context, which is $O(|G|^2 \cdot |W| \cdot |L|)$, where W is the set of all attribute values.

3.1 Lazy Classification

Suppose we have a training set (negative and positive examples wrt. some target attribute) and unclassified examples, all object descriptions given by a pattern

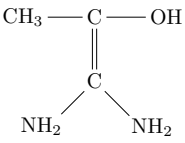
structure. Then the target attribute of the description of an object g_n to be classified with respect to the premises can be computed as the closure (w.r.t. $(G, (D*, \sqcap*), \delta*)$) of the intersection of the description of g_n with description of every object $g \in G$. If for some object g the closure contains the target attribute, g_n is classified positively by premises of $(G, (D*, \sqcap*), \delta*)$, otherwise it is classified negatively. Computationally, this can be described as the following simple two-stage procedure:

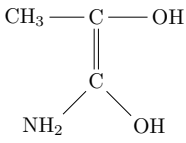
1. For every $g \in G$ compute $(\delta(g_n) \sqcap \delta(g))^\diamond$, i.e. select all objects from G whose descriptions contain $\delta(g_n) \sqcap \delta(g)$. This takes $O(|G| \cdot (p(\sqcap) + |G| \cdot p(\sqsubseteq)))$ time.

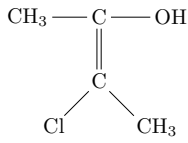
2. If for some $g \in G$ all objects from $(\delta(g_n) \sqcap \delta(g))^\diamond$ have the target attribute, classify g_n positively, otherwise negatively. This takes $O(|G|^2)$ time for looking for the target attribute in object descriptions in at most $|G|$ families of object subsets, each subset consisting of at most $|G|$ objects.

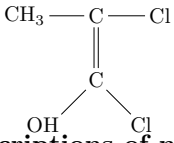
Example. Consider a training sample with four positive examples having descriptions $\{G_1\}$, $\{G_2\}$,$\{G_3\}$,$\{G_4\}$, three negative examples having descriptions $\{G_5\}$,$\{G_6\}$,$\{G_7\}$, and unclassified examples having descriptions $\{G_8\}$,$\{G_9\}$.

Descriptions of positive examples:

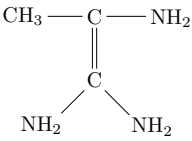
G_1 : 

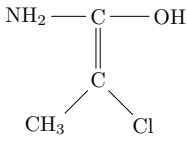
G_2 : 

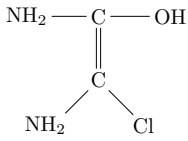
G_3 : 

G_4 : 

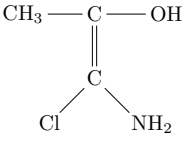
Descriptions of negative examples:

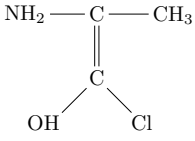
G_5 : 

G_6 : 

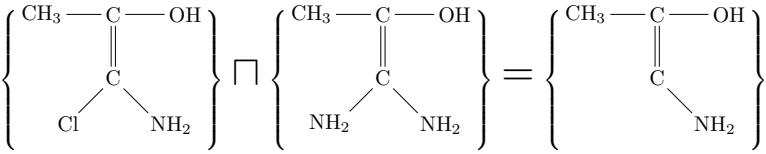
G_7 : 

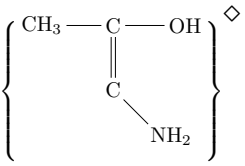
Descriptions of unclassified examples:

G_8 : 

G_9 : 

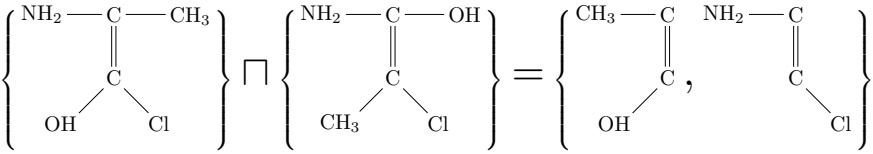
Classification results:

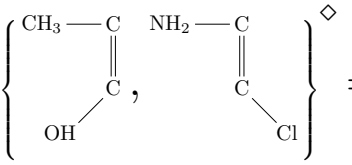




$$= \{1, 2\} \subseteq G_+,$$

i.e. example with description G_8 is classified positively.





$$= \{6\} \subseteq G_-,$$

i.e. example with description G_9 is classified negatively.

Proposition 1. *Classification of an object with premises can be done in* $O(|G| \cdot (|G| \cdot p(\sqsubseteq) + p(\sqcap)))\cdot$ *time and in* $O(|G|^2)$ *time in projections of fixed size.*

The lazy approach to classification described above is close to some standard approaches like Nearest Neighbors [37] (finding nearest classes in metric spaces), Case-Based Reasoning [20] (classifying similar to classification of similar cases), abduction in Horn theories [21] (lazy evaluation from models instead of generating implications on Boolean variables), however differs from them in being based on partially ordered structures, not metric or Boolean.

We have reduced classification to computing $(\delta(g) \sqcap \delta(g_n))^\diamond$ and testing the target attribute in all objects of this set. This computation is easily parallelizable: one partitions the dataset G in $G = G_1 \cup \ldots \cup G_k$, where k is the number of processors, computes in each G_i the set of objects $(\delta(g) \sqcap \delta(g_n))^{\diamond_i}$, tests the target attribute for all objects in the union of these sets over i. Thus, we have the following

Proposition 2. *Classification of* m *objects with premises using* k *processors can be done in* $O(|G| \cdot (|G| \cdot p(\sqsubseteq) + p(\sqcap)) \cdot m/k)$ *time and in* $O(|G|^2 \cdot n/k)$ *time in projections of fixed size.*

The computation can also be easily randomized by taking random objects from each of G_i for computing the closures. To this end, one needs to use a probabilistic relaxation of the definitions of premises and classifications.

4 Conclusion

Pattern structures propose a useful means for discovering dependencies in data given by complex ordered descriptions, such as numerical data, data given by

graphs as in chemoinformatics and natural language processing, or strings in the analysis of processes. Using projections, parallel computations and randomization, one can propose scalable approach to knowledge discovery with pattern structures by reducing algorithmic complexity from double exponential to low degree polynomial.

Acknowledgments. This work was done within the project "Mathematical models, algorithms, and software tools for intelligent analysis of structural and textual data" supported by the Basic Research Program of the National Research University Higher School of Economics (Moscow).

References

1. Arimura, H., Uno, T.: Polynomial-Delay and Polynomial-Space Algorithms for Mining Closed Sequences, Graphs, and Pictures in Accessible Set Systems. In: Proc. SDM, pp. 1087–1098 (2009)
2. Babin, M.A., Kuznetsov, S.O.: Enumeration Minimal Hypotheses and Dualizing Monotone Boolean Functions on Lattices. In: Jäschke, R. (ed.) ICFCA 2011. LNCS (LNAI), vol. 6628, pp. 42–48. Springer, Heidelberg (2011)
3. Babin, M.A., Kuznetsov, S.O.: Computing Premises of a Minimal Cover of Functional Depedencies is Intractable. Discr. Appl. Math. 161(6), 742–749 (2013)
4. Baixeries, J., Kaytoue, M., Napoli, A.: Computing Functional Dependencies with Pattern Structures. In: Proc. 9th International Conference on Concept Lattices and their Applications (CLA 2012), Malaga (2012)
5. Birkhoff, B.: Lattice Theory. ACM (1991)
6. Buzmakov, A.V., Egho, E., Jay, N., Kuznetsov, S.O., Napoli, A.: On Projections of Sequential Pattern Structures with an Application on Care Trajectories. In: Proc. 10th International Conference on Concept Lattices and their Applications (CLA 2013), La Rochelle (2013)
7. Chaudron, L., Maille, N.: Generalized Formal Concept Analysis. In: Ganter, B., Mineau, G.W. (eds.) ICCS 2000. LNCS (LNAI), vol. 1867, pp. 357–370. Springer, Heidelberg (2000)
8. Coulet, A., Domenach, F., Kaytoue, M., Napoli, A.: Using pattern structures for analyzing ontology-based annotations of biomedical data. In: Cellier, P., Distel, F., Ganter, B. (eds.) ICFCA 2013. LNCS (LNAI), vol. 7880, pp. 76–91. Springer, Heidelberg (2013)
9. Distel, F., Sertkaya, B.: On the Complexity of Enumerating Pseudo-intents. Discrete Applied Mathematics 159(6), 450–466 (2011)
10. Férré, S., Ridoux, O.: A Logical Generalization of Formal Concept Analysis. In: Ganter, B., Mineau, G.W. (eds.) ICCS 2000. LNCS (LNAI), vol. 1867, pp. 371–384. Springer, Heidelberg (2000)
11. Férré, S., King, R.D.: Finding Motifs in Protein Secondary Structure for Use in Function Prediction. Journal of Computational Biology 13(3), 719–731 (2006)
12. Finn, V.K.: Plausible Reasoning in Systems of JSM Type. Itogi Nauki i Tekhniki, Seriya Informatika 15, 54–101 (1991) (in Russian)
13. Galitsky, B.A., Kuznetsov, S.O., Samokhin, M.V.: Analyzing Conflicts with Concept-Based Learning. In: Dau, F., Mugnier, M.-L., Stumme, G. (eds.) ICCS 2005. LNCS (LNAI), vol. 3596, pp. 307–322. Springer, Heidelberg (2005)

14. Galitsky, B.A., Kuznetsov, S.O., Usikov, D.: Parse Thicket Representation for Multi-sentence Search. In: Pfeiffer, H.D., Ignatov, D.I., Poelmans, J., Gadiraju, N. (eds.) ICCS 2013. LNCS, vol. 7735, pp. 153–172. Springer, Heidelberg (2013)
15. Ganter, B., Kuznetsov, S.O.: Pattern Structures and Their Projections. In: Delugach, H.S., Stumme, G. (eds.) ICCS 2001. LNCS (LNAI), vol. 2120, pp. 129–142. Springer, Heidelberg (2001)
16. Ganter, B., Grigoriev, P.A., Kuznetsov, S.O., Samokhin, M.V.: Concept-based Data Mining with Scaled Labeled Graphs. In: Wolff, K.E., Pfeiffer, H.D., Delugach, H.S. (eds.) ICCS 2004. LNCS (LNAI), vol. 3127, pp. 94–108. Springer, Heidelberg (2004)
17. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
18. Garriga, G., Khardon, R., De Raedt, L.: Mining Closed Patterns in Relational, Graph and Network Data. Annals of Mathematics and Artificial Intelligence (2012)
19. Guigues, J.-L., Duquenne, V.: Familles minimales d'implications informatives resultant d'un tableau de donnees binaires. Math. Sci. Humaines 95, 5–8 (1986)
20. Hullermeier, E.: Case-Based Approximate Reasoning. Springer (2007)
21. Kautz, H.A., Kearns, M.J., Selman, B.: Reasoning with characteristic models. In: Proc. AAAI 1993, pp. 1–14 (1993)
22. Kaytoue, M., Kuznetsov, S.O., Napoli, A., Duplessis, S.: Mining gene expression data with pattern structures in formal concept analysis. Inf. Sci. 181(10), 1989–2001 (2011)
23. Kuznetsov, S.O.: Stability as an Estimate of the Degree of Substantiation of Hypotheses on the Basis of Operational Similarity. Nauchno-Tekhnicheskaya Informatsiya, Ser. 2 24(12), 21–29 (1990)
24. Kuznetsov, S.O.: JSM-method as a machine learning method. Itogi Nauki i Tekhniki, Ser. Informatika 15, 17–50 (1991) (in Russian)
25. Kuznetsov, S.O.: Mathematical aspects of concept analysis. J. Math. Sci. 80(2), 1654–1698 (1996)
26. Kuznetsov, S.O.: Learning of Simple Conceptual Graphs from Positive and Negative Examples. In: Żytkow, J.M., Rauch, J. (eds.) PKDD 1999. LNCS (LNAI), vol. 1704, pp. 384–391. Springer, Heidelberg (1999)
27. Kuznetsov, S.O.: Complexity of Learning in Concept Lattices from Positive and Negative Examples. Discr. Appl. Math. 142, 111–125 (2004)
28. Kuznetsov, S.O.: Computing Graph-Based Lattices from Smallest Projections. In: Wolff, K.E., Palchunov, D.E., Zagoruiko, N.G., Andelfinger, U. (eds.) KONT 2007 and KPP 2007. LNCS (LNAI), vol. 6581, pp. 35–47. Springer, Heidelberg (2011)
29. Kuznetsov, S.O., Obiedkov, S.A.: Some Decision and Counting Problems of the Duquenne-Guigues Basis of Implications. Discrete Applied Mathematics 156(11), 1994–2003 (2008)
30. Kuznetsov, S.O., Samokhin, M.V.: Learning Closed Sets of Labeled Graphs for Chemical Applications. In: Kramer, S., Pfahringer, B. (eds.) ILP 2005. LNCS (LNAI), vol. 3625, pp. 190–208. Springer, Heidelberg (2005)
31. Kuznetsov, S.O., Revenko, A.: Finding Errors in Data Tables: An FCA-based Approach. Annals of Mathematics and Artificial Intelligence (2013)
32. Liquiere, M., Sallantin, J.: Structural Machine Learning with Galois Lattice and Graphs. In: Proc. ICML 1998 (1998)
33. Luxenburger, M.: Implications partielle dans un contexte. Math. Sci. Hum (1991)

34. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient Minining of Association Rules Based on Using Closed Itemset Lattices. J. Inf. Systems 24, 25–46 (1999)
35. Ryssel, U., Distel, F., Borchmann, D.: Fast computation of proper premises. In: Proc. CLA 2011 (2011)
36. Yan, X., Han, J.: CloseGraph: Mining closed frequent graph patterns. In: Proc. KDD 2003, pp. 286–295. ACM Press, New York (2003)
37. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search - The Metric Space Approach. Springer (2006)

A Recurrent Neural Network with Non-gesture Rejection Model for Recognizing Gestures with Smartphone Sensors

Myeong-Chun Lee and Sung-Bae Cho

Dept. of Computer Science, Yonsei University
50 Yonsei-ro, Seodaemoon-gu, Seoul 120-749, Korea
lmspring@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Abstract. Gesture recognition provides a new interface to user. Various methods for the gesture recognition are feasible in smartphone environment since a number of sensors attached are gradually increasing. In this paper, we propose a gesture recognition method using smartphone accelerometer sensors. The high false-positive rate is definite if the gesture sequence data are increased. We have modified BLSTM (Bidirectional Long Short-Term Memory) recurrent neural network with non-gesture rejection model to deal with the problem. A BLSTM model classifies the input into the gesture and non-gesture classes, and the specific BLSTM models for the gestures further classify it into one of twenty gestures. 24,850 sequence data are used for the experiment, and it consists of 11,885 gesture sequences and 12,965 non-gesture sequences. The proposed method shows higher accuracy than the standard BLSTM.

Keywords: Smartphone Accelerometer Sensors, Gesture Spotting, Gesture Recognition, Recurrent Neural Network.

1 Introduction

Recently, various sensors such as accelerometer, ambient light, proximity, dual cameras, GPS, dual microphones, compass, and gyroscope are attached in a smart phone. Those not only are sophisticated but also show good performance. It is possible to recognize the user's pattern or aware the user's intention from the smartphone sensors. Especially, accelerometer is one of the most commonly used sensors for the physical movements of the user bringing the phone together. For this reason, much user interface works such as gesture recognition, and activity recognition are proposed [1].

However, there are two crucial problems when user recognition systems are developed from the smartphone sensors. One is the recognition of the non-gesture or non-activity data. The data for gesture or activity consist of meaningful and non-meaningful parts. Non-meaningful data can be half of total data or even more than gesture data. In this case, it is time-consuming to recognize both meaningful and non-meaningful data. The other is to maintain a high-accuracy rate even though the number of classes is increased. Most works about pattern recognition use machine learning methods to learn

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 40–46, 2013.

the data. However, the performance degrades when classifying a large number of classes. It is the important problem when supporting the various services to users.

In this paper, we propose a mobile gesture recognition method where data is collected from an accelerometer sensor attached to smartphone. To alleviate aforementioned two problems, a recurrent neural network called BLSTM (Bidirectional Long Short-Term Memory) is used for classifying twenty gestures. To distinguish the gesture and non-gesture, a non-gesture rejection model is proposed.

2 Related Works

The relevant research about gesture recognition is divided into several groups. One is the vision-based gesture recognition. Q. Chen et al. proposed a hand gesture recognition method with the combination of statistical and syntactic analyses. They used a web camera as the input device and constructed the system for real time environment [2]. Tran et al. suggested a 3-D posture tracking for real time hand gestures. They used nearest neighbor clustering algorithm based on LCS (Longest Common Subsequence) similarity measure of joint angle dynamics [3]. Next group of methods is to recognize gestures based on EMG (Electromyogram) sensor, which is a device to measure the electrical potentials generated by muscle cells. Zhang et al. used both accelerometer and EMG to classify 72 Chinese Sign Language (CSL) [4]. Wheeler et al. utilized the EMG to recognize their arm movement to control a joystick. The Bayesian decomposition method is proposed to distinguish individual muscle groups with the goal of enhancing gesture recognition [5].

Another method recognizes gestures based on accelerometer. It is divided into two groups again, using accelerometer chip only and using the accelerometer attached in a smartphone. Xu et al. used MEMS 3-axes accelerometer for seven hand gestures with the Hopfield network as associative memory for classification algorithm. A decision tree and multi-stream HMMs (Hidden Markov Models) were utilized as decision-level fusion to get classification result [6]. Each method has merits and demerits. In case of the vision-based approach, it is possible to recognize the multiple objects. However, it has a limitation to record data because of the constraints of the camera placement. The detailed data can be collected from EMG but the data cannot be collected wirelessly. Accelerometer is more suitable device to collect the gesture data since it is used without the constraints of place and can be collected the gesture data accurately. For our problem, we use an accelerometer in a smartphone. The related works using accelerometer are shown in Table 1.

Table 1. Related Workd using Mobile Accelerometer

Author	Year	Classifier	Gestures	Collector	Overview
Liu *et al.* [7]	2009	DTW	8	Smartphone	Personalized gesture using uWave
Niezen *et al.* [8]	2009	HMM, ANN, DTW	8	Smartphone	Performance comparison between many classifiers

Table 1. (*Continued*)

Author	Year	Classifier	Gestures	Collector	Overview
Min *et al.* [9]	2010	DTW, NB, K-means	20	Smartphone	DTW model selection through NB
Marasovic *et al.* [10]	2011	K-NN	7	Smartphone	Combination of the PCA and K-NN
Akl *et al.* [11]	2011	DTW, AP	18	Wii mote	Dimensionality reduction through RP

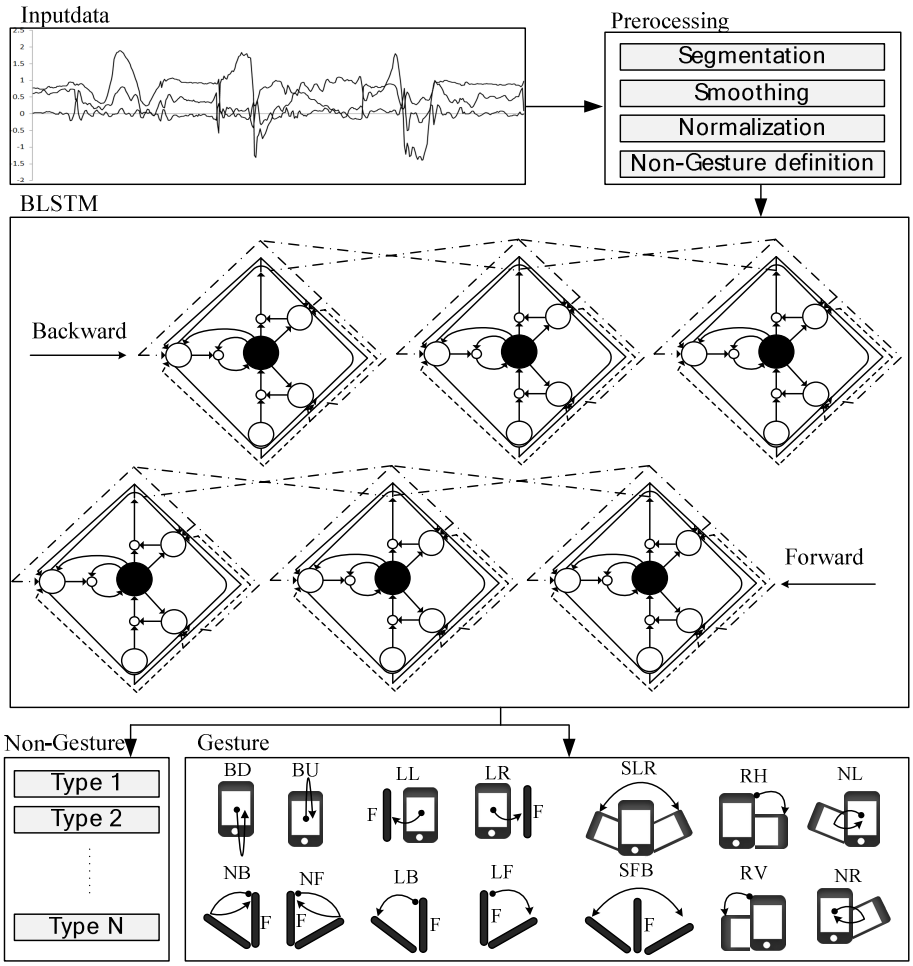


Fig. 1. Overview of the proposed method

3 The Proposed Method

This paper aims to enhance the accuracy by using non-gesture rejection model. The entire system configuration is shown in Fig 1. The accelerometer data collected from a smart phone are segmented by using sliding window and average variation. The preprocessed data are classified after training. We adopt the recurrent neural network based on BLSTM which is a hybridization of BRNN (bidirectional recurrent neural network) and LSTM. A set of training data is used to classify the gestures and non-gestures, and the other set of training data for classifying the twenty gesture classes.

LSTM is an extension of the recurrent neural network. It uses the three gates that can store and access the data collected from the rest of the network. The gates are activated from logistic sigmoid activation function. The performance outweighs HMM used for time-series classification problem. The Hyperbolic tangent activation function is used for squashing functions. The basic calculation of each gate is the same with standard artificial neural network [12].

The input gate determines whether the input values put the memory cell or not.

$$\alpha_\iota^t = \sum_{i=1}^{I} w_{i\iota} x_i^t + \sum_{h=1}^{H} w_{h\iota} \beta_h^{t-1} + \sum_{c=1}^{C} w_{c\iota} s_c^{t-1}$$
$$\beta_\iota^t = f(\alpha_\iota^t) \tag{1}$$

where α_ι^t is a state of the input gate at time t. It is calculated from input values, the output of other networks, and state of the memory cell. I, H, and C mean the numbers of input node, hidden node, and cell, respectively. w is the weight of connected nodes. f(x) is the logistic sigmoid function to activate the input gate.

The output gate determines whether the information is output or not. The calculation of the output gate is similar with the input gate.

$$\alpha_\Phi^t = \sum_{i=1}^{I} w_{i\Phi} x_i^t + \sum_{h=1}^{H} w_{h\Phi} \beta_h^{t-1} + \sum_{c=1}^{C} w_{c\Phi} s_c^{t-1}$$
$$\beta_\Phi^t = f(\alpha_\Phi^t) \tag{2}$$

where α_Φ^t is a state of the forget gate at time t and β_Φ^t is a state after applying the activation function.

Eq. (3) is a calculation that is generated by forget gate, the state of cell, the state of input gate, and state of α_c^t after applying hyperbolic tangent activation function. Note that the fixed weight value 1.0 is used for preserving the information in the memory cell.

$$\alpha_c^t = \sum_{i=1}^{I} w_{ic} x_i^t + \sum_{h=1}^{H} w_{hc} \beta_h^{t-1}$$
$$S_c^t = \beta_\Phi^t S_c^{t-1} + \beta_\iota^t g(\alpha_c^t) \tag{3}$$

Forget gate provides the information to reset the memory cell.

$$\alpha_\omega^t = \sum_{i=1}^{I} w_{i\omega} \alpha x_i^t + \sum_{h=1}^{H} w_{h\omega} \beta_h^{t-1} + \sum_{c=1}^{C} w_{c\omega} s_c^{t-1}$$
$$\beta_\omega^t = f(\alpha_\omega^t) \tag{4}$$

where ω is the output gate and β_{ω}^{t} is the state of an output gate after applying the activation function at time t.

Eq. (5) is the definition of the cell output. To activate the cells, hyperbolic tangent is used and multiplies the state of output gate.

$$\beta_{c}^{t} = \beta_{\omega}^{t} \mathrm{e}(s_{c}^{t}) \tag{5}$$

For training the LSTM recurrent neural network, we use the Back Propagation Through Time (BPTT) algorithm.

4 Experimental Results

For the experiments, Samsung Omnia smartphone is used with MS Windows Mobile 6.1 as a platform. The acceleration is sampled at 50Hz. 30 people of 10~60 years old participate in the experiments. The collected data are divided into generations and date. Total amount of the data consists of 11,885 gesture sequences and 12,965 non-gesture sequences. The number of files which is used in the experiment is 1,075.

Table 2 shows the symbol of twenty gestures, direction and meaning. Rotating and tilting to hold their physical states after the movement. Tapping represents the hand or finger stroke on a smart phone surface. In the case of shaking, subjects shake the devices two or more times in a specific direction. Snapping has an angular acceleration while bouncing moves straightly to a direction and reflected back where both are the kind of pendulum movement.

Table 2. Types of Gesture Classes

Symbol	Direction	Meaning	Symbol	Direction	Meaning
NL	Left	Snapping	TL	Left	Tapping
NR	Right		TR	Right	
NF	Forward		TF	Forward	
NB	Backward		TB	Backward	
BU	Up	Bouncing	TT	Top	
BD	Down		TM	Bottom	
LL	Left	Tilting	RH	Horizontal	Rotating
LR	Right		RV	Vertical	
LF	Forward		SLR	Left-Right	Shaking
LB	Backward		SFB	Forward-Backward	

For the first experiment, the data are divided into a ratio of seven to three for training and test, respectively. The 17,470 sequences are used for training, and 7,380 sequences are used for test. Each sequence is distributed randomly. The results through all experiments in this work are compared with standard BLSTM. Average accuracy rate of the proposed method is 91.11% and standard BLSTM is 89.17%. For the second experiment, we group the data as generation. 18,490 sequences are used for training and approximately 2,100 sequences are used for testing each generation. Fig 2 shows that the proposed method outperforms standard BLSTM for all generations.

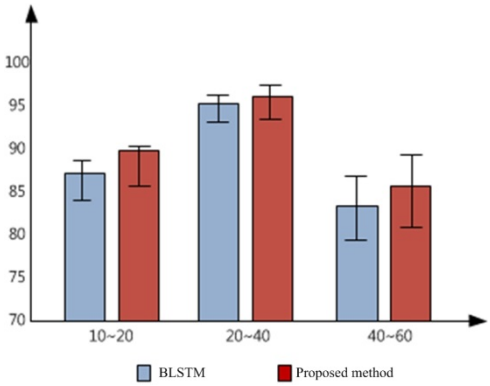


Fig. 2. The generational experiment

5 Concluding Remarks

In this paper, we have collected the accelerometer data from a smartphone and classified the data by using a recurrent neural network with non-gesture rejection model. Total classes of the data are twenty one including non-gestures. The performance of the standard BLSTM compared with the proposed method and our approach outperforms the standard BLSTM. However, it is not satisfactory to apply in the real world. It might be possible to obtain higher accuracy if the data are grouped with the similar meaning because some gestures have similar features. For a future work, we will investigate the characteristics of features for similar gestures.

References

1. Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Cambell, A.T.: A survey of mobile phone sensing. IEEE Communications Magazine 48(9), 140–150 (2010)
2. Chen, Q., Georganas, N.D., Petriu, E.M.: Hand gesture recognition using haar-like featuresand a stochastic context-free grammar. IEEE Trans. on Instrumentation and Measurement 57(8), 1562–1571 (2008)
3. Tran, C., Trivedi, M.M.: 3-D posture and gesture recognition for interactivity in smart spaces. IEEE Trans. on Industrial Informatics 8(1), 178–187 (2012)
4. Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., Yang, J.: A framework for hand gesture recognition based on accelerometer and EMG sensors. IEEE Trans. on Systems, Man, and Cybernetics, Part A: Systems and Humans 41(6), 1064–1076 (2011)
5. Wheeler, K.R., Chang, M.H., Knuth, K.H.: Gesture-based control and EMG decomposition. IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 36(4), 503–514 (2006)
6. Xu, R., Zhou, S., Li, W.J.: MEMS accelerometer based nonspecific-user hand gesture recognition. IEEE Sensors Journal 12(5), 1166–1173 (2012)
7. Liu, J., Wang, Z., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: accelerometer-based personalized gesture recognition and its applications. IEEE Int. Conf. on Pervasive Computing and Communications, 1–9 (2009)

8. Niezen, G., Hancke, G.P.: Evaluating and optimising accelerometer-based gesture recognition techniques for mobile devices. AFRICON, 1–6 (2009)
9. Min, J.-K., Choe, B.-W., Cho, S.-B.: A selective template matching algorithm for short and intuitive gesture UI of accelerometer-builtin mobile phones. In: Cong. on Nature and Biologically Inspired Computing, pp. 660–665 (2010)
10. Marasovic, T., Papic, V.: Accelerometer-Based Gesture Classification Using Principal Component Analysis. In: Int. Conf. on Software, Telecommunications and Computer Networks, pp. 1–5 (2011)
11. Akl, A., Feng, C., Valaee, S.: A novel accelerometer-based gesture recognition system. IEEE Trans. on Signal Processing 59(12), 6197–6205 (2011)
12. Hochreiter, S., Schmidhuer, J.: Long short-term memory. Neural Computation 9(8), 1735–1780 (1997)

Granular Attribute Selection: A Case Study of Rough Set Approach to MRI Segmentation

Sebastian Widz[1] and Dominik Ślęzak[2]

[1] Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warsaw, Poland
[2] Institute of Mathematics, University of Warsaw
ul. Banacha 2, 02-097 Warsaw, Poland
sebastian.widz@ibspan.waw.pl, slezak@mimuw.edu.pl

Abstract. We introduce a new rough set inspired approach to attribute selection. We consider decision systems with attributes specified by means of two layers: 1) general meta-attribute descriptions, and 2) their specific realizations obtained by setting up parameters of procedures calculating attribute values. We adopt methods designed for finding rough set reducts within the sets of attributes grouped into clusters, where each cluster contains potentially infinite amount of attributes realizing a single meta-attribute. As a case study, we discuss a rough set framework for multi-spectral Magnetic Resonance Image (MRI) segmentation.

Keywords: Rough set reducts, Attribute hierarchies, MRI segmentation.

1 Introduction

Attribute selection is an important step in the process of constructing knowledge representation and classification models [1]. In rough sets, this step corresponds to excluding irrelevant attributes and, as a result, simplifying descriptions of objects in information systems and decision tables [2]. There are numerous approaches to searching for rough set reducts – the sets of attributes containing no irrelevant elements [3]. Rough set attribute selection and classification techniques can be also embedded into hierarchical systems that utilize domain knowledge to approximate complex concepts [4].

In [5], the clustering methods are used to identify groups of attributes that can replace each other in reducts. By working with such attribute granules, one can make the reduct derivation process faster and more intuitive. In this paper, we note that attributes can be grouped also with respect to other criteria. Quite often, domain experts sketch some high level attribute descriptions, which then need to get mathematically modeled. One could actually say that each of such meta-attributes corresponds to a granule of attributes which are derived using different parameters from the same model.

As a case study, we discuss the problem of MRI analysis [6], which has been already investigated in the rough set literature from the perspectives of both image segmentation [7] and tissue classification [8]. We outline several examples of meta-attributes corresponding to different ways of looking at the MRI slices. We show how to run the process of attribute selection given a huge space of attribute parameterizations, and how to avoid multiple realizations of the same meta-attributes within single reducts.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 47–52, 2013.

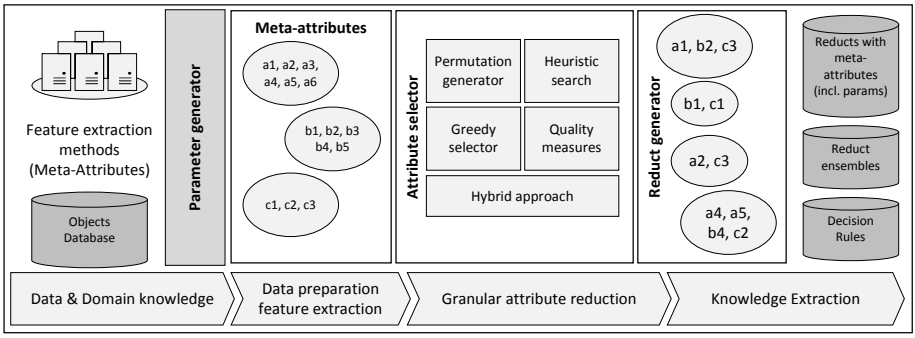


Fig. 1. Granular attribute selection framework

2 Rough Set Reduct Construction

Let us use standard notation to represent data [2]. By a decision system we mean a tuple $\mathbb{A} = (U, A \cup \{d\})$, where U is a set of objects, A is a set of attributes and $d \notin A$ is a distinguished decision attribute. We treat attributes $a \in A$ as functions $a : U \to V_a$. We say that objects $u, u' \in U$ are discerned by $a \in A$, iff $a(u)$ is different from $a(u')$. We say that $B \subseteq A$ is a decision superreduct, denoted by $B \in \mathcal{S}$, $\mathcal{S} \subseteq \mathcal{P}(A)$, iff each pair of objects discerned by d is also discerned by an element of B. We say that $B \in \mathcal{S}$ is a decision reduct, iff there are no proper subsets of B belonging to $\mathcal{S}$.

In practice, the criteria for being a decision superreduct can be formulated in many other ways adjusted to the types of attributes, expected level of noise in data, and so on [8]. Generally, we should work with families of approximate decision superreducts $\mathcal{S}^* \subseteq \mathcal{P}(A)$ satisfying the following two conditions: 1) $\mathcal{S} \subseteq \mathcal{S}^*$; 2) if $B \in \mathcal{S}^*$ and $B \subseteq C$, then $C \in \mathcal{S}^*$. The first condition means that standard decision superreducts should be always a special case of any of approximate superreduct extensions. The second condition means that the considered criteria are monotonic [9].

Superreduct criteria can be related to degrees of information that subsets of attributes bring about the decision. Let us consider function $I_d : \mathcal{P}(A) \to \mathcal{I}$, where $\langle \mathcal{I}, \preceq \rangle$ is a partially ordered space of information degrees. We assume that if $B \subseteq C$, then $I_d(B) \preceq I_d(C)$. We can think about approximate superreducts corresponding to certain levels of information, i.e., families $\mathcal{S}^* \subseteq \mathcal{P}(A)$ such that if $B \in \mathcal{S}^*$ and $I_d(B) \preceq I_d(C)$, then $C \in \mathcal{S}^*$. Surely, for large data sets it is important to work with functions I_d that are easy to compute. It is also useful to develop procedures for computing values of $I_d(B)$ for many subsets $B \subseteq A$ in the same time [10].

For a given function $I_d : \mathcal{P}(A) \to \mathcal{I}$ and a corresponding family $\mathcal{S}^* \subseteq \mathcal{P}(A)$, the idea of defining decision reducts is the same as before. We say that $B \in \mathcal{S}^*$ is an approximate decision reduct, iff there are no proper subsets of B belonging to $\mathcal{S}^*$. Fast derivation of approximate decision reducts containing possibly small amounts of attributes is highly important for representing and utilizing data dependencies. However, for decision systems with huge number of attributes, approximate decision reducts are very hard to compute. Also, it is then difficult to interpret which approximate dependencies between attributes and decisions may be meaningful for users.

Algorithm 1. Superreduct generator.

Input: meta-attributes $\{C_1, ..., C_k\}$
Output: decision superreduct $B \in \mathcal{S}^*$

```
1  B = ∅; i = 1; fail = FALSE;
2  while B ∉ S* or fail do
3      fail = TRUE;
4      select m > 0 random parameter settings for C_i;
5      I_max = I_d(∅);
6      a_max = NULL;
7      for j = 1, m do
8          generate attribute c_{i_j} ∈ C_i using the j-th parameter setting;
9          if I_d(B ∪ {c_{i_j}}) ≠ I_max then
10             I_max = I_d(B ∪ {c_{i_j}});
11             a_max = c_{i_j};
12             fail = FALSE;
13         end
14     end
15     if a_max is not NULL then
16         B = B ∪ {a_max};
17         a_max = NULL;
18     end
19     i = (1 − i/k) ∗ i + 1;
20 end
21 return B;
```

Let us consider a scenario where attributes correspond to the outputs of parameterized feature extraction methods applied for some complex measurements. Such methods may generate different attributes depending on selected parameters. On the other hand, such attributes have analogous meaning for users. Reducts containing too many of such attributes might be considered as ambiguous. Figure 1 presents a framework for grouping conceptually analogous attributes into clusters and ensuring that calculated reducts do not contain multiple elements of the same clusters too often. Users can then focus on general data dependencies more easily. We call such clusters meta-attributes. Each cluster refers to a feature extraction method and its parameter generator. Thus, each meta-attribute can be represented by one or many "real" attributes.

Algorithm 1 shows how to construct an ordering of attributes corresponding to the input meta-attributes. It is a random-greedy approach based on our earlier research [5]. For each meta-attribute, we select several parameter settings and calculate values of the corresponding attributes. An attribute bringing the highest information gain in terms of a given information function $I_d : \mathcal{P}(A) \to \mathcal{I}$ is selected. We do not assume creation of a complete decision system. We calculate values of only those attributes which are selected by the algorithm. The resulting $B \in \mathcal{S}^*$, together with an ordering of added attributes, can be passed to a typical permutation based algorithm for reduct derivation [8]. However, sometimes $B \in \mathcal{S}^*$ cannot be obtained. In such cases, one should revisit expectations with respect to a reachable level of decision information.

Table 1. Examples of MRI meta-attributes: SOM (self-organizing map), HC (histogram clustering), EDG (edge detection filter), NBR (neighborhood filter), MSK (image mask filter)

	Description	**Parameters**
SOM	Performs unsupervised image segmentation	Number of modalities; Number of clusters; Learning rate; Number of iterations; Radius
HC	Performs unsupervised image segmentation based on the image histogram	Number of clusters; Bucket size; Minimum cluster distance; Peak separability; Approximation Degree; Number of clusters
EDG	Detects boundaries between tissues	Noise Threshold
NBR_{SOM}	For boundary regions assigns major SOM tissue class label.	Window shape; Window size
NBR_{HC}	For boundary regions assigns major HC tissue class label	Window shape; Window size
MSK	Encodes voxel position relative to brain central point	Number of mask regions; Region radius size; Background threshold

3 Rough Set Inspired MRI Segmentation

As a case study, let us present a rough set based approach to human brain MRI segmentation [6]. Segmentation is the process of assigning class labels to data containing spatially varying information [7]. As an example, we can consider the Simulated Brain Database (SBD) containing 3-D volumetric multi-spectral MRI images in three modalities (T1, T2, PD).[1] In this case, the goal is to classify each voxel within a brain image slice to one of tissue classes such as: cerebrospinal fluid (CSF), gray matter (GM) and white matter (WM), background (BCG), as well as others (skin, bone, fat).

In order to define meta-attributes, we can use several feature extraction methods. Combined with specific parameters, each of those methods generates an attribute labeling voxels with specific values. At the conceptual level, each meta-attribute provides some domain related information about the analyzed images. On the other hand, feature extraction parameters are often very technical, not necessarily understandable by medical experts. We need to separate those two realities and provide tools allowing users to analyze attribute dependencies without worrying about any technical settings.

The primary source of information is the voxel magnitude value registered in three MRI modalities. Using this source, we can construct more advanced attributes. In our research, we employed two unsupervised clustering algorithms: a self-organizing map (SOM) and our own method for image histogram discretization [8]. We also utilized the open source Insight Segmentation and Registration Toolkit (ITK)[2] providing a library of image filters. One can employ such filters for edge detection, voxel neighborhood analysis and relative voxel positioning against a brain center. Each of considered feature extraction methods can be applied separately to all image modalities and can be considered as a meta-attribute describing a single modality. Table 1 summarizes meta-attributes and possible parameters used within our segmentation framework.

[1] `www.bic.mni.mcgill.ca/brainweb`
[2] `www.itk.org`

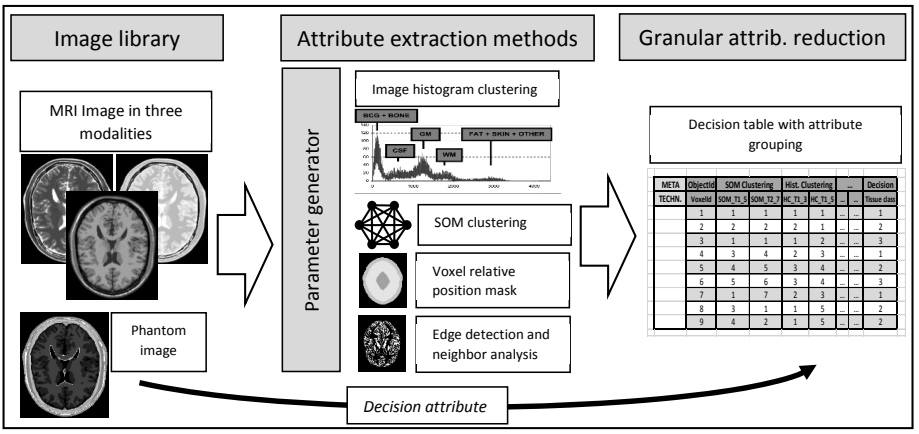


Fig. 2. Application of granular attribute selection framework in MRI image segmentation

Figure 2 illustrates the resulting procedure. Let us discuss how it differs from our earlier approach described in [8]. First of all, as already mentioned in the previous section, we do not create a decision system containing all attributes and their values. Instead, we calculate only those attributes which seem to be sufficiently relevant according to our heuristic attribute selection procedure outlined by Algorithm 1.

Furthermore, some of considered attributes correspond to simple classifiers (e.g., SOM) that need to be trained in order to generate values for particular voxels-objects. In our previous studies, we trained each of such classifiers-attributes separately, searching for parameters providing the best outcomes. However, such an approach might lead to over-training, with no ability to search for possibly coarser realizations of particular meta-attributes that could work well together within a dynamically constructed decision reduct. Actually, there is a strong analogy between this way of looking at reducts and a well known methodology of working with ensembles of weak local classifiers [1]. In case of framework outlined in this paper, we do not train such local classifiers at all. Instead, we assume that the attribute selection and reduction mechanisms will set up combinations of parameters for all meta-attributes within a single procedure.

Last but not least, let us emphasize that in our approach we do not store the trained attributes-classifiers but only their parameter settings. Then, for a new image, we run the related feature extraction method with the previously learnt parameters. Therefore, it is important to remember that parameter settings corresponding to attributes selected by Algorithm 1 should not be applied to the previously unseen images too straightforwardly. Such an approach might be suitable for images of comparable characteristics. However, if images are expected to vary, we should rather attempt to express parameter values as corresponding to some local image properties. For example, a noise threshold used for realization of the EDG meta-attribute can be expressed in terms of the mean magnitude computed over all voxels in a given image. This way, our MRI segmentation model should be more stable and resistant with respect noises and anomalies.

4 Conclusions

We presented a new rough set inspired approach to attribute selection, for decision systems created using parameterized attribute value generation procedures. We described two levels of formation of attributes: meta-attributes serving as conceptual descriptions of general feature extraction methods, as well as their specific realizations constitute an actual decision system. As a case study, we presented an extension of our previous MRI image segmentation framework. In future, we will investigate usefulness of our approach also in other areas, such as predictive analytics based on sensor measurements and knowledge discovery based on text reports.

Acknowledgments. This research was supported by grant O ROB/0010/03/001 ("Modern Engineering Tools for Decision Support for Commanders of the State Fire Service of Poland during Fire & Rescue Operations in the Buildings") founded by Polish National Centre for Research and Development, as well as grants 2011/01/B/ST6/03867 and 2012/05/B/ST6/03215 founded by Polish National Science Centre.

References

1. Kuncheva, L.I., Diez, J.J.R., Plumpton, C.O., Linden, D.E.J., Johnston, S.J.: Random Subspace Ensembles for fMRI Classification. IEEE Transactions on Medical Imaging 29(2), 531–542 (2010)
2. Pawlak, Z.: Rough Sets – Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers (1991)
3. Kruczyk, M., Baltzer, N., Mieczkowski, J., Dramiński, M., Koronacki, J., Komorowski, J.: Random Reducts: A Monte Carlo Rough Set-based Method for Feature Selection in Large Datasets. Fundamenta Informaticae 127(1-4), 273–288 (2013)
4. Nguyen, S.H., Bazan, J., Skowron, A., Nguyen, H.S.: Layered Learning for Concept Synthesis. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B.z., Swiniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 187–208. Springer, Heidelberg (2004)
5. Janusz, A., Ślęzak, D.: Rough Set Methods for Attribute Clustering and Selection. Applied Artificial Intelligence (2014)
6. Koo, J.J., Evans, A.C., Gross, W.J.: 3-D Brain MRI Tissue Classification on FPGAs. IEEE Transactions on Image Processing 18(12), 2735–2746 (2009)
7. Maji, P., Pal, S.K.: Maximum Class Separability for Rough-Fuzzy C-Means Based Brain MR Image Segmentation. In: Peters, J.F., Skowron, A., Rybiński, H. (eds.) Transactions on Rough Sets IX. LNCS, vol. 5390, pp. 114–134. Springer, Heidelberg (2008)
8. Widz, S., Ślęzak, D.: Approximation Degrees in Decision Reduct-based MRI Segmentation. In: Howard, D., Rhee, P.K. (eds.) Frontiers in the Convergence of Bioscience and Information Technologies 2007, FBIT 2007, Jeju Island, Korea, October 11-13, pp. 431–436. IEEE Computer Society (2007)
9. Yao, Y., Zhao, Y., Wang, J.: On Reduct Construction Algorithms. In: Gavrilova, M.L., Tan, C.J.K., Wang, Y., Yao, Y., Wang, G. (eds.) Transactions on Computational Science II. LNCS, vol. 5150, pp. 100–117. Springer, Heidelberg (2008)
10. Kowalski, M., Stawicki, S.: SQL-Based Heuristics for Selected KDD Tasks over Large Data Sets. In: Ganzha, M., Maciaszek, L.A., Paprzycki, M. (eds.) Federated Conference on Computer Science and Information Systems, FedCSIS 2012, September 9-12, pp. 303–310. IEEE, Wrocław (2012)

Performance Tuning of PCA by CFS-Shapley Ensemble and Its Application to Medical Diagnosis

S. Sasikala[1], S. Appavu Alias Balamurugan[2], and S. Geetha[3]

[1] Anna university, Tamil Nadu, India
[2] K.L.N. College of Information Technology, Tamil Nadu, India
[3] Thiagarajar College of Engineering, Tamil Nadu, India
nithilannsasikala@yahoo.co.in, app_s@yahoo.com, sgeetha@tce.edu

Abstract. Selection of optimal features is an important area of research in medical data mining systems. Principal component analysis (PCA) is one among the most popular feature selection methods. Still PCA faces a drawback – i.e., the measurements from all of the original features are used in the projection to the lower dimensional space. Hence this work is aimed to tune the performance of PCA and classify the medical profiles. The proposed method is realized as an ensemble procedure with three steps – (i) feature selection using PCA, (ii) feature ranking with CFS and (iii) dimension reduction using Shapley Values Analysis. The variance coverage parameter of PCA is adjusted so as to yield maximum accuracy which are measured with specificity, sensitivity, precision and recall. This facilitates the selection of a compact set of superior features with uncompromised detection rates, remarkably at a low cost. To appraise the success of the proposed method, experiments were conducted across 6 different medical data sets using J48 decision tree classifier, which showed that the proposed procedure improves the classification efficiency and accuracy compared with individual usage.

Keywords: Data mining, Dimensionality reduction, Feature Extraction, Feature selection, Principal component analysis, Shapley value Analysis, Classification.

1 Introduction

In the classification tasks on the medical datasets, researchers notice that it is common that a considerable number of features are not informative because they are either irrelevant or redundant with respect to the class concept. Ideally, we would like to use the features which have high separability power while ignore or pay less attention to the rest. A suitable feature set can simplify both the pattern representation and the classifiers consequently; the resulting classifier will be more efficient. Identifying the optimal feature subset according to classification performance is referred to as feature selection process.

Feature selection methods are typically divided into two types. The peculiarity is made depending on whether feature subset evaluation is performed using the learning

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 53–59, 2013.

algorithm employed in the classifier design (i.e., wrapper) or not (i.e., filter). Filter based approaches are computationally more efficient than wrapper based approaches since they evaluate the goodness of selected features using criteria that can be tested quickly. Nevertheless this could lead to the selection of non-optimal features, particularly, when the features are found to be dependent on the classifier. This may restrain the classifier performance and make it to perform poorly. Contrarily, the wrapper methods perform evaluation by training the classifier using a validation set. Even though this is a slower procedure, the method guarantees selection of more optimal features for the classifier employed.

Feature reduction has been applied to several areas in medicine [1], [2]. These are e.g. classification of ultrasound liver tissues using wavelet transform , classification of EEG signals for operating brain-computer interfaces , detection of mass lesions in digital mammograms , classification of hepatic lesions from computed tomography images , segmenting digital chest radiographs , etc. Huang et al. [3] predicts type 2 diabetic patients by employing a feature selection technique as supervised model construction to rank the important attributes affecting diabetes control. Ingui and Rogers [4] derive and validate an automated method to determine the set of patients for whom a given primary care physician holds overall clinical responsibility.

The restrictions of existing research for features subset selection clearly recommend the use of an improved framework for feature selection that facilitates efficient analysis of both feature relevance and redundancy for medical data. Therefore, a hybrid feature selection method that ensembles both filter and wrapper models of feature subset selection is proposed in this paper. The ultimate motive of the proposed method is to enhance performance of the classifier thereby assuring more accurate diagnosis in a relatively short span of time with reduced complexity. This will assist the doctors in selecting and prescribing the most suitable treatment for the patient and in quick recovery.

2 Proposed PCA-CFS-Shapley Values Ensemble Feature Selection Algorithm

2.1 Feature Extraction Using PCA

Principal Component Analysis (PCA) is a numerical technique used for extracting information from a multi-variety dataset. This process is performed via having principal components of original variables with linear combinations identified. While the original dataset with the maximum variability is represented with the first principal component, the dataset from the remaining with the maximum variability is represented with these second principal components. The process goes on consecutively as such, with the dataset from the remaining with the maximum variability being represented with the next principal component. While μ represents the number of all principal components, and $\eth$ represents the number of the significant principal components among all principal components. It is clear there in that $\eth \leq \mu$. Hence, PCA is employed as a data-reducing technique and the core

features in the chosen medical datasets are identified and subjected to further processing.

2.2 Ranking Using Correlation Based Feature Selection

Correlation-based Feature Selection (CFS) is a method that evaluates subsets of attributes rather than individual attributes. The heuristic in equation (1) assigns high scores to subsets containing attributes that are highly correlated with the class and have low score for inter correlation with each other. The equation that assesses the importance of a feature is given as *score* in (1)

$$score(S) = \frac{Kr_{feature-class}}{\sqrt{K + K(K-1)\alpha r_{feature-feature}}} \tag{1}$$

Where ' S' denotes the subset of features that are selected currently and 'K' denotes the number of features in subset S.The correlation measures are '$r_{feature-class}$' is the average feature-class correlation and '$r_{feature-feature}$' is the average feature-feature correlation. 'α' denotes a scaling factor (usually α = 0:25)

2.3 Feature Selection Using Shapley Values Analysis

The contribution of each variable to the classification task is estimated based on the Shapley value[5].The Shapley value is defined below .Let the marginal importance of variable i to a variable subset S , with $i \notin S$, be

$$\Delta_i(s) = v(s \cup \{i\}) - v(s) \tag{2}$$

Here function v associates with every non-empty subset S of F , a real number $v(s)$ (the value of S) with $v(\{\varphi\}) = 0$. M is the number of the variable in variable sets F with all the variables. The unbiased d-bounded estimator for the Shapley value is given as

$$\varphi_i(v) = \frac{1}{|\Pi_d|} \sum_{x \in \Pi_d} \Delta_i\left(S_i(\pi)\right) \tag{3}$$

where π_d is the set of sampled permutations on subsets of size d .

2.4 Hybrid Feature Selection Algorithm - CFS-SA Ensemble

A hybrid feature subset selection strategy that ensembles both filter and wrapper models of feature subset selection named as CFS- Shapley Values Analysis Ensemble is proposed in this paper. In the first stage, the filter model is used to rank the features

by CFS algorithm between each feature and its class. The feature's ranking processing is as follows:

a) Calculate the Score metric of variable set as defined in equation (1).

b) Find the β - the average of $Score(S)$

c) Rank the $Score(S)$ in descending order.

d) Returning the top β features with the top β weight.

S is the selected sub set of features and $Score(S)$ represents the score values of the set S. The β value is chosen by average weight in which the feature is ranked by their weight value in descending order.

In this stage, the best β features are selected, which are highly relevant to the classes however, the federation contribution of the features in a feature subset need not be found. Actually, the coalitional contribution of the features is important for the targeted classification tasks. Therefore, the wrapper model based feature selection algorithm is completed in the second stage by employing Shapley value based feature analysis and selection is described as follows:

a) Let the top selected features β form the feature set, SF.

b) For each feature $f \in SF$, calculate the d-bounded estimated contribution value (Shapley value), Sv.

c) If max $Sv \prec \delta$, eliminate λ features which are with the λ lowest Shapley values.

d) Return SF.

d is the d-bounded used to estimate contribution values. δ is a contribution value threshold. F is the set of all the feature of a data set, each feature $f \in F$. λ is the number of features eliminated in each phase.The function v in equation (3) can be calculated as follows:

a) Obtaining the new train data, $Trn.X$ and $Trn.Y$, in the new feature subset space.

b) Generating a classifier from the training set, using $Trn.X$, $Trn.Y$

c) Classifying the validation set data, $Vali.X$.

d) $$v(s) := \frac{|\{x | f_s(x) = y, (x, y) \in Validation|}{|Validation|}, x \in Vali.X, y \in Vali.y \quad (4)$$

e) Returning $v(S)$.

3 Performance Analysis

In order to show the effectiveness of the proposed method, we have applied the proposed method (PCA-CFS-SA ensemble) across 6 different medical datasets

including binary and multiclass problems like namely Cardiac Arrhythmia , Dermatology, Hepatitis, Pima Diabetes, Pregnancy Back Ache and Biomed datasets available with the UCI Machine Learning repository [6]. The datasets are represented using triplet form (a,b,c) where 'a' denotes the total instances in the dataset, 'b' denotes the number of features and 'c' denotes the number of classes available in the dataset. The classification performance of each application was evaluated using the statistical parameters like sensitivity, specificity, precision and recall by tuning the variance coverage factor in the feature extraction process. The proposed method is realised with J48 classifier that have been implemented in Weka environment [7]. The 10-fold cross validation procedure is adapted for the classification stages. All experiments were run on a machine with a 2.8GHz CPU,4GB of RAM, 500 GB HDD space, and the Windows7 operating system. Figure 1 portrays the number of features selected out of different sequence of operations and it could be easily noted that the proposed system offers the fewest number of features in all the cases. These set of features also help to improve the sensitivity and specificity of the classifiers. Due to want of space, only the best results obtained are reported here. J48 classifier outperformed the other classifiers. Table 1 reports the best variance coverage factor at which the maximum sensitivity, specificity, precision and recall values achieved while building the classifier model. It could be noted that the proposed PCA-CFS-SA ensemble algorithm produced significant accurate diagnosis with J48 classifier.

Table 1. Performance Evaluation of the proposed system on various Medical Datasets

Medical Data Set	**Observation Parameters**	**LDA**	**Information Gain (IG)**	**PCA**	**PCA + CFS**	**PCA + SA**	**Proposed PCA + CFS + SA Ensemble**
Cardiac Arrhythmia (452,280 16)	Variance Coverage	-	-	0.95	0.95	0.95	0.95
	Precision	0.523	0.58	0.506	0.522	0.538	0.8756
	Recall/ Sensitivity	0.613	0.628	0.615	0.628	0.6215	0.9215
	Specificity	0.532	0.628	0.582	0.591	0.5865	0.8865
	Running Time (sec)	1.56	0.06	0.09	0.15	1.8	2.83
Dermatology (366, 35, 6)	Variance Coverage	-	-	0.65	0.65	0.65	0.65
	Precision	0.932	0.955	0.947	0.955	0.951	0.979
	Recall/ Sensitivity	0.922	0.954	0.945	0.954	0.9495	0.9795
	Specificity	0.932	0.954	0.947	0.955	0.951	0.981
	Running Time (sec)	0.23	0.01	0	0.11	1.02	1.56

Table 1. (*Continued*)

Hepatitis (155,20,2)	Variance Coverage	-	-	0.85	0.85	0.85	0.85
	Precision	0.759	0.797	0.797	0.832	0.813	0.849
	Recall/ Sensitivity	0.787	0.819	0.787	0.819	0.803	0.833
	Specificity	0.712	0.819	0.739	0.797	0.768	0.798
	Running Time (sec)	0.08	0.03	0	0.09	0.15	0.18
Pima Diabetes (768,9,2)	Variance Coverage	-	-	0.9	0.9	0.9	0.9
	Precision	0.724	0.735	0.734	0.749	0.735	0.743
	Recall/ Sensitivity	0.726	0.738	0.74	0.749	0.749	0.745
	Specificity	0.724	0.738	0.734	0.744	0.739	0.769
	Running Time (sec)	0.05	0.01	0.02	0.05	0.15	1.2
Pregnancy Back Ache (180,33,2)	Variance Coverage	-	-	0.7	0.7	0.7	0.7
	Precision	0.712	0.722	0.742	0.742	0.739	0.756
	Recall/Sensitiv ity	0.814	0.816	0.833	0.861	0.847	0.877
	Specificity	0.812	0.816	0..833	0.861	0.7515	0.7815
	Running Time (sec)	0.04	0.01	0	0.03	0.11	0.12
Biomed (209,9,2)	Variance Coverage	-	-	0.6	0.65	0.65	0.65
	Precision	0.894	0.899	0.899	0.912	0.911	9.325
	Recall/ Sensitivity	0.895	0.900	0.900	0.914	0.907	0.937
	Specificity	0.895	0.900	0.900	0.961	0.93	0.96
	Running Time (sec)	0.01	0.02	0.05	0.09	1.23	1.65

4 Conclusion

We presented an improved and accurate feature selection algorithm that is based on PCA. The proposed method can be used for variable selection in medical diagnosis fulfilling the objectives aimed: very small sets of non-redundant features are obtained while preserving predictive accuracy. However the suitability of the proposed algorithm on high dimensional data sets is not tested. As a part of our future work, we are planning to apply it for high dimensional data sets. Additionally, improving the

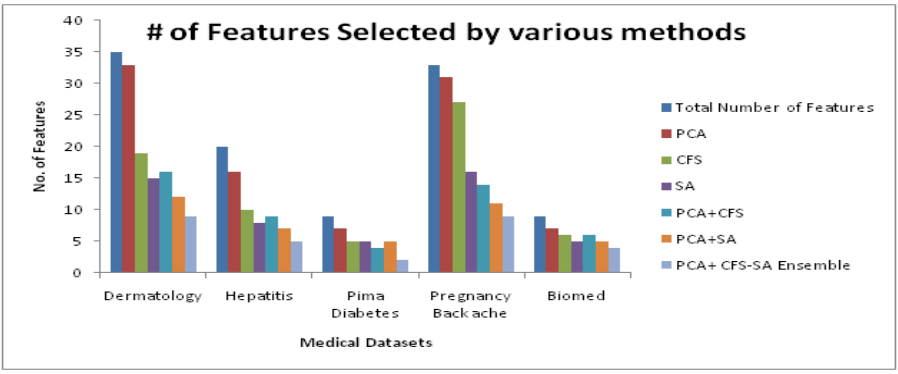


Fig. 1. Number of features selected by various feature selection algorithms

speed of the algorithm remains to be a future goal. With all these factors addressed, the proposed PCA, CFS and Shapley Values Analysis ensemble should probably become part of the "standard tool-box" of feature selection methods for medical data analysis.

Acknowledgments. This work is supported in part by the University Grant Commission Major Research Project under grant no. **F.No.:39-899/2010 (SR).**

References

1. Fernández-Navarro, F., et al.: Evolutionary Generalized Radial Basis Function neural networks for improving prediction accuracy in gene classification using feature selection. Applied Soft Computing Journal (2012), doi:10.1016/j.asoc.2012.01.008
2. Li, D.-C., Liu, C.-W., Hu, S.C.: A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. Artificial Intelligence in Medicine 52(1), 45–56 (2011)
3. Huang, Y., McCullagh, P., Black, N., Harper, R.: Feature selection and classification model construction on type 2 diabetic patients data. Artificial Intelligencte Medicine Journal 41, 251–262 (2007)
4. Ingui, B.J., Rogers, M.A.: Searching for clinical prediction rules in MEDLINE. Journal of American Medical Information Association 8(4), 391–397 (2007)
5. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games Annals of Mathematics Studies, vol. II(28), pp. 307–317. Princeton University Press, Princeton (1953)
6. Hettich, S., Blake, C., Merz, C.: UCI repository of machine learning databases (1998), `http://www.ics.uci.edu/mlearn/MLRepository.html`
7. Weka 3: Machine Learning Software in Java. The University of Waikato software documentation, `http://www.cs.waikato.ac.nz/_ml/weka`

A Modified *K*-Modes Clustering Algorithm

Partha Sarathi Bishnu and Vandana Bhattacherjee

Deptartment of Computer Science & Enggineering, Birla Institute of Technology, Ranchi, Jharkhand, 834001, India
psbishnu@gmail.com, bhattacherjeev@yahoo.co.in

Abstract. The main aim of this paper is to apply a new dissimilarity measure, and handle boundary data properly in K-Modes clustering thereby increasing the clustering efficiency. Moreover our proposed algorithm identifies the outlier data efficiently.

Keywords: Data mining, clustering, K-Modes algorithm, outlier.

1 Introduction

To overcome the drawback of partitional clustering techniques like K-Means, K-Medoid algorithms which are popular clustering algorithms for numeric only datasets, Huang, [1] suggested K-Modes clustering algorithm to handle categorical data. Software cost estimation, bioinformatics, and computer networks are some of the areas of application for the K-Modes clustering algorithm [4]. There are some issues which directly affect the quality of the clusters using K-Modes algorithm: first, the initial selection of the cluster centers [1], [2], [6], second, the similarity or dissimilarity measures between the two non numeric data [3], [4], [5], third, proper handling of the boundary data [7] and finally outlier detection. In this paper our main objectives are threefold; first, we introduce a new dissimilarity metric, second, we introduce a new technique to handle boundary data properly and third, identification of the outlier data to handle the second, third, and fourth issue respectively. We have compared our technique with three existing techniques suggested by Huang, 1998 (K-Modes 1)[1], Huang and Ng, 1999, (K-Modes 2) [7] and He et al., 2005 (K-Modes 3) [3] with help of real datasets from UCI machine learning repository.

The outline of this paper is as follows: in section two we describe our proposed dissimilarity measure, and a method to handle boundary data and outliers. In section three we explain the experiments which we carried out. In section four we present the results and this is followed by conclusions in section five.

2 Proposed *K*-Modes Clustering Algorithm

2.1 *K*-Modes Clustering Algorithm

First we discuss the simple K-Modes clustering algorithm for the sake of completeness [1], [4]. A categorical dataset (O) is defined by a set (D) of attributes

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 60–66, 2013.

$I_1, I_2, ..., I_d$where d is the dimension of the dataset. Each attribute I_e is described by a set of categorical values $DOM(I_e)$ which are finite and unordered. For $a, b \in DOM(I_e)$, either $a = b$ or $a \neq b$. A data $o_i \in O$, where $i = 1, 2, ...n$ and n is the number of data, can be represented as a conjunction of attribute value pairs $[I_1 = a_1] \wedge [I_2 = a_2] \wedge ... \wedge [I_d = a_d]$, where $a_e \in DOM(I_e)$. We use ε to represent the missing value. Moreover, if $o_i, o_j \in O$, and let $o_i = [a_1, a_2, ..., a_d]$ and $o_j = [b_1, b_2, ..., b_d]$ then we write $o_i = o_j$ if $a_e = b_e$, for $1 \leq e \leq d$ and the relation $o_i = o_j$ does not mean that they are the same data, but rather that the two data have equal values on attributes (I_e), where $1 \leq e \leq d$ [4].

Let $o_i, o_j \in O$, be two categorical data represented by $[a_1, a_2, ..., a_d]$ and $[b_1, b_2, ..., b_d]$, respectively. The distance (simple matching dissimilarity) between o_i and o_j is defined as $dis(o_i, o_j) = \sum_{e=1}^{d} \delta(a_e, b_e)$ where

$$\delta(a_e, b_e) = \begin{cases} 0 & : \quad a_e = b_e \\ 1 & : \quad a_e \neq b_e \end{cases} \tag{1}$$

In K-Modes the objective of clustering categorical dataset into $K(<< n)$ clusters is to search W and M that minimize the cost function $F(W, M)$, where $F(W, M) = \sum_{k=1}^{K} \sum_{i=1}^{n} w_{ki} dis(m_k, o_i)$ (2) , subject to $w_{ki} \in \{0, 1\}$, $1 \leq k \leq K$ and $1 \leq i \leq n$ (3), $\sum_{k=1}^{K} w_{ki} = 1$, $1 \leq i \leq n$ (4) and , $0 < \sum_{i=1}^{n} w_{ki} < n$, $1 \leq k \leq K$ (5), where $W = [w_{ki}]$ is a k-by-n $\{0, 1\}$ matrix and $m_k \in M$, $1 \leq k \leq K$, is the k^{th} cluster mode.

2.2 New Dissimilarity Measure

We introduce a new dissimilarity measure ($dist$) for the K-Modes clustering algorithm so as to minimize the cost function $F_{BB}(W, M)$, where $F_{BB}(W, M) = \sum_{k=1}^{K} \sum_{i=1}^{n} w_{ik} dist(m_k, o_i)$ subject to conditions as in (3), (4), and (5).

Definition 1. *Let $o_i, o_j \in O$ be two categorical data represented by $[a_1, a_2, ..., a_d]$ and $[b_1, b_2, ..., b_d]$, respectively. Let $dist(o_i, o_j)$ be the distance between two objects. The distance is defined as follows: $dist(o_i, o_j) = 1 - \frac{\alpha_{ij}}{\beta_{ij}}$, where $\alpha_{ij} = \sum_{e=1}^{d} \chi(a_e, b_e)$, where,*

$$\chi(a_e, b_e) = \begin{cases} 1 & : \quad a_e = b_e \\ 0 & : \quad a_e \neq b_e \end{cases} \tag{6}$$

and, $\beta_{ij} = 2d - \alpha_{ij}$, where d is the dimension of the data. Here, $\frac{\alpha_{ij}}{\beta_{ij}}$ measures the similarity between two objects as the degree of relative overlap between the two objects. α_{ij} counts the number of attributes for which the values in o_i, o_j match while β_{ij} counts those which do not.

Theorem 1. *Let M be fixed and consider the problem: $min_w F_{BB}(W, M)$, subject to (3) to (5). The minimizer M is given by*

$$w_{ki} = \begin{cases} 1, & : \quad dist(m_k, o_i) \leq dist(m_h, o_i), 1 \leq h \leq K, \\ 0, & : \quad otherwise. \end{cases}$$

Theorem 2. *Let* $m_k = [m_{k1}, m_{k2}, ..., m_{kd}]$ *be the mode of the* k^{th} *cluster and the domain* ζ_{a_j} *of attributes* a_j *be* $\{a_j^1, a_j^2, ..., a_j^{n_j}\}$, $1 \leq j \leq d$ *and* n_j *is the cardinality of domain of* a_j. *Denote arbitrary object* o_i *by* $[a_{i1}, a_{i2}, ..., a_{id}]$. *Then* $F_{BB}(W, M) = \sum_{k=1}^{K} \sum_{i=1}^{n} w_{ki} dist(m_k, o_i)$ *is minimized if and only if* $m_{kj} = a_j^r$, *where,* $a_j^r \in \zeta_{a_j}$ *satisfies:* $|\{w_{ki} | o_{ij} = a_j^r, w_{ki} = 1\}| \geq |\{w_{ki} | o_{ij} = a_j^t, w_{ki} = 1\}|$, $1 \leq t \leq n_j$, $1 \leq j \leq d$.

Proof. For a given W, $F_{BB}(W, M) = \sum_{k=1}^{K} \sum_{i=1}^{n} w_{ki} dis(m_k, o_i) = \sum_{k=1}^{K} \psi_k$. Note that all the inner sums ψ_k of $F_{BB}(W, M)$ are non negative and independent. Then minimizing $F_{BB}(W, M)$ is equivalent to minimizing each inner sum. By definition 1, when $m_{kj} = a_j^t$ we have $\psi_k = \sum_{i=1}^{n} w_{ki} dist(m_k, o_i) = \sum_{i=1}^{n} w_{ki}(1 - \frac{\alpha_{ki}}{\beta_{ki}}) = N_k - \sum_{i=1}^{n} \frac{\alpha_{ki}}{\beta_{ki}})$, where N_k is the number of data in k^{th} cluster. Since N_k is nonnegative, ψ_k is minimized if $\frac{\alpha_{ki}}{\beta_{ki}}$ is maximized for $1 \leq i \leq n$. Since $\frac{\alpha_{ki}}{\beta_{ki}}$ is the similarity (or degree of relative overlap) between mode m_k and object o_i, it is maximized when $t = r$ i.e., $m_{kj} = a_j^r$. The result follows. □

2.3 Proper Handling of Boundary Objects

The data o_i should be assigned to cluster number k_q if $dis(o_i, m_k) < dis(o_i, m_l)$, $1 \leq l, q \leq K$ and $l \neq q$. Now if $dis(o_i, m_k) = dis(o_i, m_l)$, then o_i is termed as boundary data [7]. It is conventional to assign o_i to cluster number k_q or k_l, which ever comes first. However, in this paper we suggest a novel approach to assign proper cluster number for a boundary data. To handle the proper assignment of boundary data we use one of the criteria, these being the average distance ($AvgDist$), total distance ($Dist$), and size (of q^{th} cluster, c_q) (n_q). The distances are computed as $AvgDist_{iq} = \frac{Dist_{iq}}{n_q}$, where $Dist_{iq} = \sum_{j=1}^{n_q} dist(o_i, o_{qj})$, $o_{qj} \in c_q$. The boundary data a_i should be assigned to q^{th} cluster if $AvgDist_{iq} < AvgDist_{il}$, $1 \leq l, q \leq K$, $l \neq q$. Further, in a rare situation when $AvgDist_{iq} = AvgDist_{il}$, data o_i should be assigned to q^{th} cluster if $Dist_{iq} > Dist_{il}$ or if $n_q > n_l$. In the rarest of situation where all these values happen to coincide and the data is not detected as outlier then we can assign it randomly to any of q or l.

2.4 Outlier Detection

An outlier is a data which is far away (very different) from the rest of the data.

Definition 2. *Let* $o_i = [a_{i1}, a_{i2}, ..., a_{id}]$ *and* $o_j = [b_{j1}, b_{j2}, ..., b_{jd}]$ *be two data objects, where* $1 \leq j \leq n, i \neq j, o_i, o_j \in O$. *Then* o_i *is outlier data if* $a_{ie} \neq b_{je}$, $1 \leq j \leq n, 1 \leq e \leq d$, $i \neq j$.

Since none of the attributes of an outlier data o_i matches with any attribute of other data, neither simple technique nor boundary data assignment technique will help to assign cluster number and data will be floating among clusters. So we assign that data point o_i as outlier data and remove it from the dataset.

Definition 3. *Let* $o_i = [a_{i1}, a_{i2}, ..., a_{id}]$ *then* o_i *is outlier data if* $AvgDist_{ij} = 1$, *for all* $1 \leq j \leq K$.

Example 1. Let cluster 1 consist of three data $[a, a, b; a, a, c; a, a, d]$, where mode of cluster 1 is $m_1 = [a, a, b]$ and cluster 2 consists of four data $[a, a, e; a, c, b; a, c, b; a, a, f]$ where mode of cluster 2 is $m_2 = [a, a, b]$ (assume modes are same by chance). Consider another data $o_p = [a, c, d]$, which is not part of any cluster and is a boundary data because of the same distance $dist(o_p, m_1) = dist(o_p, m_2) = 1 - \frac{1}{2*3-1} = 0.8$ from the mode 1 and mode 2. To assign the cluster number we calculate the $AvgDist_{p1}(0.7)$, and $AvgDist_{p2}(0.65)$. Here, $AvgDist_{p2} < AvgDist_{p1}$, so data $[a, c, d]$ should be assigned to second cluster. Furthermore the data $o_i = [x, y, z]$ (initially denoted as border data) is the outlier data because no attribute of o_i is matching any attribute of with any other data and $AvgDist_{i1} = AvgDist_{i2} = 1$.

2.5 Proposed K-Modes Clustering Algorithm

Next we give a formal algorithm to explain our proposed algorithm (Algorithm 1) as follows:

```
Algorithm 1: Proposed K-Modes Algorithm
Input: O, K
Output: C (clusters)
Step1: randomly select initial modes from given dataset O;
Step2: (re) assign cluster number to each data by calculating
the distance (dissimilarity) between the data and the modes;
Step3: identify and label boundary data;
Step4: handle each boundary data separately to assign the cluster
number and identify the outlier data and remove the outlier data
from the dataset;(from 2^nd iteration outlier data will not appear)
Step5: (re)identify the modes;
Step6: repeat steps 2 to 5 till convergence criteria is satisfied;
```

The time complexity of the step 1 is $O(1)$ and steps 2 and 5 is $O(ntKd)$, where t is the number of iteration. The time complexity to handle boundary data and outlier data is $O(xntd)$, where x ($x \ll n$) is the number of boundary data and $O(ynd)$where y is the number of outlier data and it is identified in first iteration only i.e. when $t = 1$ (step 4) respectively.

3 Experiments

3.1 Experimental Analysis

To evaluate our proposed algorithm we have used four real datasets from UCI machine learning repository (http://archive.ics.uci.edu/ml/), namely Lung cancer data (LD), Breast cancer data (BD), Zoo data (ZD), and Soybean data (SD).

The number of data are 32, 699, 101, and 47 respectively. The dimensions are 57, 11, 18, 35 respectively. We eliminate attributes or data with any missing values. No real datasets consist of any outlier data, so we have used two synthetic datasets (syn1 and syn2) consisting of 100 and 500 data respectively and the dimensions are 10 and 15 respectively. We randomly add five outlier data to each synthetic dataset. More over to show the scalability of our proposed algorithm we execute all the algorithms on different dimensions and sizes (Figure 1a and Figure 1b). To achieve our first objective of introducing a new dissimilarity measure for categorical data, we execute all but step 3 and step 4 of our proposed algorithm (algorithm 1), and name this as K-Modes 4 i.e., we are not handling boundary and outlier data in K-Modes 4. Finally, we execute all our steps with new dissimilarity measure as well as handling of boundary and outlier data and name it as K-Modes 5 in Table 1. For all the data sets LD, BD, ZD and SD the number of clusters are as per the number of classes present i.e., 3, 2, 7, and 4 respectively. All the algorithms were executed 100 times and the best result of accuracy was reported. All the algorithms iterate 30 times (we set this value as convergence criteria). Efficiency is expressed in terms of execution time. To evaluate the efficiency of the K-Modes clustering algorithms we use three evaluation parameters namely, $accuracy = \frac{\sum_{i=1}^{K} m_i}{n}$, $precision = \frac{\sum_{i=1}^{K} \frac{m_i}{n_i}}{K}$, and $recall = \frac{\sum_{i=1}^{K} \frac{m_i}{m_i+p_i}}{K}$. Where, m_i is the number of data that are correctly assigned to i^{th} cluster, n_i is the total number of data present in the i^{th} cluster, p_i is the data incorrectly rejected from the i^{th} cluster [4], [7]. The experiments were conducted on a PC with an Intel Celeron processor, (1.30 GHz), and 256 MB RAM running the Windows XP operating system. All the K-Modes algorithms (K-Modes 1 [1], K-Modes 2 [7], K-Modes 3 [3], K-Modes 4 and K-Modes 5 (proposed)) have been coded in Octave -3.2.4.

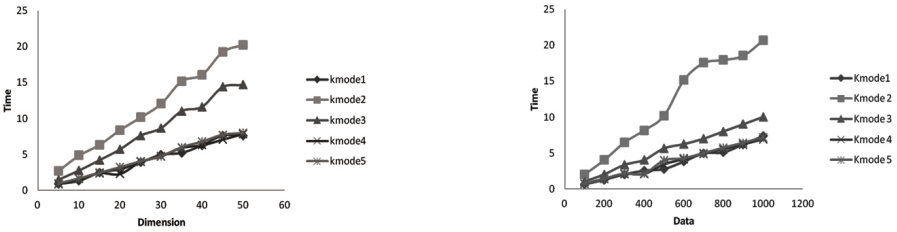
Table 1. Experiments on Real Datasets

Dataset	Evaluation	K-Modes 1	K-Modes 2	K-Modes 3	K-Modes 4	K-Modes 5
Lung Cancer	Accuracy	0.9687	0.9687	0.9687	0.9687	1
	Precision	0.9666	0.9666	0.9761	0.9761	1
	Recall	0.9743	0.9743	0.9629	0.9629	1
	Time	3.5001	8.9062	8.4531	3.5469	4.4219
Breast Cancer	Accuracy	0.9399	0.9799	0.8446	0.9399	0.9985
	Precision	0.9580	0.9725	0.9038	0.9580	0.9989
	Recall	0.9128	0.9847	0.7738	0.9128	0.9765
	Time	8.6875	19.047	15.344	8.8901	19.3570
Zoo	Accuracy	0.8514	0.8811	0.8514	0.9108	0.9306
	Precision	0.8464	0.8623	0.8397	0.9191	0.9285
	Recall	0.9644	0.8507	0.8642	1	0.9642
	Time	7.3125	35.344	10.469	8.1406	19.500
Soybean	Accuracy	1	1	1	1	1
	Precision	1	1	1	1	1
	Recall	1	1	1	1	1
	Time	4.5938	13.0611	11.0942	3.7188	4.9219

3.2 Results

We have evaluated the clustering efficiency of our algorithms with other K-Modes algorithms. The performance of K-Modes 5 algorithm in terms of accuracy and

precision is the best for all datasets and recall is the second best only for zoo data. The performance of K-Modes 4 is second best to K-Modes 5 and better than all the other algorithms.To evaluate the scalability performance of our proposed algorithm we compare execution time with other existing algorithms. For breast cancer and zoo data the execution time of K-Modes 5 is high but it is reasonably low for the other two datasets. From the plot (Figure 1a and Figure 1b) it is seen that the times for our algorithm K-Modes 4 and K-Modes 5 fall in the lowest category (using synthetic datasets). It is also seen that the algorithms are scalable in terms of dimension and size of the data. Moreover our proposed algorithm (K-Modes 5) detected correctly all the outlier data from the synthetic datasets syn1 and syn2, and hence our algorithm shows the potentiality to identify the outlier data.



(a) Dimension versus Time (b) Data versus Time

Fig. 1. Scalability Performance

4 Conclusions

In this paper we have suggested a new dissimilarity measure and techniques for handling border data and outlier data to increase the K-Modes clustering quality. The results show the effectiveness of our proposed K-Modes algorithm by comparing with other popular existing K-Modes algorithms upon real as well as synthetic datasets.

References

1. Huang, Z.: Extensions to the K-Means algorithm for clustering large data sets with categorical values. Data Mining Knowledge Discovery 2(3), 283–304 (1998)
2. Cao, F., Liang, J., Bai, L.: A new initialization method for categorical data clustering. Expert Systems with Applications 36, 10223–10228 (2009)
3. He, Z., Deng, S., Xu, X.-F.: Improving K-Modes algorithm considering the frequencies of attribute values in mode. In: Hao, Y., Liu, J., Wang, Y.-P., Cheung, Y.-M., Yin, H., Jiao, L., Ma, J., Jiao, Y.-C. (eds.) CIS 2005. LNCS (LNAI), vol. 3801, pp. 157–162. Springer, Heidelberg (2005)
4. Ng, M.K., Li, M.J., Huang, J.Z., He, Z.: On the impact of dissimilarity measure in K-Modes clustering algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 503–507 (2007)

5. Cao, F., Liang, J., Li, D., Bai, L., Dang, C.: A dissimilarity measure for the K-Modes clustering algorithm. Knowledge-Based Systems 26, 120–127 (2012)
6. Bai, L., Liang, J., Dang, C.: An initialization method to simultaneously find initial cluster centers and the number of clusters for clustering categorical data. Knowledge-Based System 24, 785–795 (2011)
7. Huang, Z.X., Ng, M.K.: A fuzzy k-modes algorithm for clustering categorical data. IEEE Transactions on Fuzzy Systems 7(4), 446–452 (1999)

BiAS: A Theme Metric to Model Mutual Association

Ramkishore Bhattacharyya

Microsoft Corporation, Redmond, WA - 98052
ramb@microsoft.com

Abstract. Identifying likeness between events is one of the fundamental necessities in machine learning and data mining techniques. Though grouping of events usually happens on their proximity in Euclidean space or the degree of similarity or the extent of linear dependence, certain applications like keyword and document clustering, phylogenetic profiling and feature selection tend to yield better results if events are grouped based on their mutual association. This paper presents a metric, the Bidirectional Association Similarity (*BiAS*) to quantify the degree of mutual association between a pair of events. We put forward generalized formulation to compute *BiAS* and establish unidirectional correspondence with the Jaccard and the cosine similarities. The measure can be suitably incorporated with clustering algorithms in grouping mutually associative events with adding precision to the discovered knowledge.

Keywords: Bi-directional association similarity, BiAS, clustering, cosine similarity, Jaccard index, mutual association.

1 Introduction

Identifying similarity or dissimilarity between events is one of the fundamental necessities in machine learning, data mining techniques which is measurable through computation of certain statistical metrics e.g. the Euclidean distance (*E*), the Jaccard (*J*) [4] [5] and cosine similarity (*C*) or the Pearson correlation coefficient (*P*). Though grouping of events, in general, happens on their proximity in Euclidean space or degree of similarity or extent of linear dependence, certain applications like document clustering, phylogenetic profiling, feature selection etc. tend to yield better results if events are grouped on the basis of pair-wise mutual association. A pair of events g_i and g_j is mutually associative if the probability of occurrence of g_j is high given the occurrence of g_i and vice-versa. Unfortunately, neither similarity nor correlation guarantees mutual associativity as they do not have individual control over $p(g_j/g_i)$ or $p(g_i/g_j)$, p denotes probability. Two events g_i and g_j can be similar or correlated if,

- there exist unidirectional association of the form $g_i \rightarrow g_j$ or
- there really exist mutual association of the form $g_i \rightarrow g_j$ and $g_j \rightarrow g_i$ or
- there exist a third event g_k such that $g_k \rightarrow g_i$ and $g_k \rightarrow g_j$ hold simultaneously

So existence and directionality of association remain vague unless explicitly verified.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 67–73, 2013.

The *Bidirectional Association Similarity* (*BiAS*) models mutual associativity with the help of two simultaneous associations $g_i \rightarrow g_j$ and $g_j \rightarrow g_i$ [1] and is quantified by $\beta(g_i,g_j) = p(g_j/g_i)*p(g_i/g_j)$. Given two pre-specified thresholds μ and τ, g_i and g_j are mutually associative if both of $p(g_j/g_i)$, $p(g_i/g_j) \geq \mu$ and $\beta(g_i,g_j) \geq \tau$. We prove subsequently that mutually associative events are also similar, but similarity does not guarantee mutual associativity. Hence, *BiAS* adds both way associativity constraints on top of similarity. It helps in pruning out loosely coupled expression vectors with adding precision to discovered knowledge. The salient contributions in this paper are:

- Concept and foundation of *BiAS* on the basis of mutual associativity
- Formulation of *J* as a function of $p(g_j/g_i)$, $p(g_i/g_j)$ and deriving its lower bound
- Generalized formulation of *BiAS* for real-valued attributes
- Bridging connections between the lower bounds of *BiAS*, *J* and *cos*
- Finally, proving capability of *BiAS* to be integrated with clustering algorithms

2 Mutual Association and Generalized Jaccard Index

Let us assume $G = \{g_1, g_2, \ldots, g_n\}$ be a set of n events where each of g_i is a d-dimensional Boolean vector. For any g_i, a '1' (or '0') at l^{th} dimension indicates its presence (or absence) in l^{th} experiment. Further, let $T(g_i)$ be a set of integers $j \in \{1, 2, \ldots, d\}$, so that $g_{ij} = 1$. Hence, $T(g_i) \cap T(g_j)$ is a set of integers m, where both $g_{im} = 1$ and $g_{jm} = 1$. Also, assume that $c(g_i) = |T(g_i)|$, the frequency of occurrence of the event g_i in the dataset. For preciseness, we denote $p(g_j/g_i)$ as μ_f and $p(g_i/g_j)$ as μ_b. The generalized form of *Jaccard* index *J* for a set of events G' can be formulated as:

$$J(G') = \left|\cap_{g \in G'} T(g)\right| \Big/ \left|\cup_{g \in G'} T(g)\right| .$$

Eqn. (1) in the following states a key relationship between *J*, μ_f and μ_b. It can be easily proven by replacing μ_f with $|T(g_i) \cap T(g_j)| / |T(g_i)|$ and μ_b with $|T(g_i) \cap T(g_j)| / |T(g_j)|$.

$$J(g_i, g_j) = \frac{1}{(1/\mu_f) + (1/\mu_b) - 1} \tag{1}$$

Given μ_f, μ_b and the constraints, eqn. (1) can be utilized as the criterion function to be minimized through non-linear programming optimization to yield a lower bound on *J*.

Lemma 1. Let $\mu_f, \mu_b \geq \mu$ and $\mu_f * \mu_b \geq \tau$. The *Jaccard* index of two events g_i and g_j,

$$J(g_i, g_j) \geq \frac{\mu\tau}{\mu^2 + \tau - \mu\tau} \tag{2}$$

Proof. The inequality is proven via the Karush-Kuhn-Tucker (KKT) theorem.

The Karush-Kuhn-Tucker (KKT) theorem. If the function f(x) has a minimum at x* in the feasible set and if $\nabla f(x^*)$ and $\nabla g_i(x^*)$, i = 1,2,…,m, exist ('∇' denotes partial derivative with respect to all x_i), then there exist an m-dimensional vector λ such that

- $\nabla f(x^*) + \Sigma_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0$, for i = 1, 2,…, m.
- $g_i(x^*) \bullet 0$, for i = 1, 2,…, m.

- $\lambda_i [g_i(x^*) - b_i] = 0$, for i = 1, 2,…, m.
- $\lambda_i \geq 0$, for i = 1, 2,…, m.

(x*, λ) is called a KKT point, λ is the Dual Vector or the Lagrange Multiplier.

Minimizing eqn. (1) is equivalent to maximizing $(1/\mu_f) + (1/\mu_b) - 1$ which is again equivalent to minimizing $1 - (1/\mu_f) - (1/\mu_b)$. Let us now formulate the problem: Minimize: $f(\mu_f, \mu_b)$: $1 - (1/\mu_f) - (1/\mu_b)$…(i), subjected to: g_1: $\mu_f \bullet \mu \Rightarrow -\mu_f \bullet -\mu$, g_2: $\mu_b \bullet \mu \Rightarrow -\mu_b \bullet -\mu$ and g_3: $\mu_f * \mu_b \bullet \tau \Rightarrow -\mu_f * \mu_b \bullet -\tau$. Clearly, $\nabla f(\mu_f, \mu_b)$ and $\nabla g_i(\mu_f, \mu_b)$, for i = 1, 2, 3 exist. Hence, by KKT theorem, $(1/\mu_f^2) - \lambda_1 - \lambda_3 * \mu_b = 0$ … (ii) and $(1/\mu_b^2) - \lambda_2 - \lambda_3 * \mu_f = 0$ … (iii). Also, $-\lambda_1(\mu_f - \mu) = 0$… (iv), $-\lambda_2(\mu_b - \mu) = 0$… (v), $-\lambda_3(\mu_f * \mu_b - \tau) = 0$… (vi) and $\lambda_i \bullet 0$, for i = 1, 2, 3… (vii).

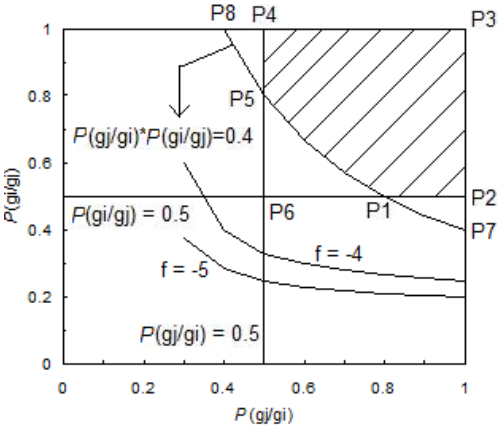


Fig. 1. Plot for eqns. μ_f=0.5, μ_b=0.5, $\mu_f * \mu_b$=0.4 and f, shaded area shows solution region

Without any constraint on μ_f and μ_b, $0 \bullet \mu_f, \mu_b \bullet 1$. Different eqns. viz. $\mu_f = p(g_j/g_i) = \mu$, $\mu_b = p(g_i/g_j) = \mu$ and $\mu_f * \mu_b = p(g_j/g_i) * p(g_i/g_j) = \tau$ (for $\mu = 0.5$ and $\tau = 0.4$) are plotted in Fig. 1. The objective function (eqn. (i)) is also plotted for f = -4 and f = -5. As can be noticed, f is minimized in the direction toward the center (0,0). Also, the feasible region for μ_f and μ_b is shaded in Fig. 1. Clearly, this is a convex region as all the points on the line segment joining any two points from the region completely remain within it. The regions inside P1P2P7 and P4P5P8 get excluded from the solution due to the constraint on $p(g_j/g_i)$ or $p(g_i/g_j)$. The constraint on $p(g_j/g_i) * p(g_i/g_j)$ excludes the region P1P5P6 from the solution. Together, they ensure high confidence in one direction with a minimal in the reverse. It is evident from eqn. (i) that both of μ_f and μ_b need to be minimized for minimization of f. Hence the solution must be somewhere on the line segment P1P5.

Consider Point P1. P1 is the point of intersection of $\mu_b = \mu$ and $\mu_f * \mu_b = \tau$. At P1, $\mu_f \neq \mu$. Hence, from eqn. (iv), $\lambda_1 = 0$. Given this, eqn. (ii) yields $\lambda_3 = 1/(\mu_f^2.\mu_b)$. Replacing λ_3 with its value in eqn. (iii), $\lambda_2 = (\mu_f - \mu_b) / (\mu_f.\mu_b^2) = (\tau - \mu^2) / (\mu_f^2.\mu^2)$ (replacing μ_b by μ and μ_f in the numerator by τ/μ_b). Since, $\tau \bullet \mu^2$, $(\tau - \mu^2) \bullet 0$. **Thus, all** λ_1, λ_2 and λ_3 are

• 0, which satisfies eqn. (vii). Putting $\mu_b = \mu$ and $\mu_f = \tau / \mu$ in the objective function (in (i)) f = 1 – (μ / τ) – (1 / μ) … (viii).

Consider Point P5. Similarly, on point P5, it can be shown that all λ_1, λ_2 and λ_3 are • 0 and value of the objective function (in (i)) f = 1 – (μ / τ) – (1 / μ), same as (viii).

Consider Any Point on the Line Segment P1P5. Here, μ_f, $\mu_b \neq \mu$ but $\mu_f * \mu_b = \tau$. From eqn. (iv), (v) and (vi), $\lambda_1 = 0$ and $\lambda_2 = 0$. Putting this in eqn. (ii) and (iii), $\lambda_3 = 1/\mu_f^2.\mu_b = 1/\mu_f.\mu_b^2 \Rightarrow \mu_f = \mu_b = \tau^{0.5}$. Thus, all λ_1, λ_2 and λ_3 are • 0, which satisfies eqn. (vii) and f = 1 – (1 / $\tau^{0.5}$) – (1 / $\tau^{0.5}$) = 1 – (2 / $\tau^{0.5}$)… (ix).
(viii) – (ix) = (2/$\tau^{0.5}$) – (μ/τ) – (1/μ) = – (τ + μ^2 – 2 . $\mu\tau^{0.5}$) / $\mu\tau$ = – ($\tau^{0.5}$ – μ)^2 / $\mu\tau$ • 0. Hence, (viii) • (ix) and optimum solution exist at P1 or P5. For P1, $\mu_f = \tau/\mu$ and $\mu_b = \mu$, for P5 $\mu_f = \mu$ and $\mu_b = \tau/\mu$. Putting, the values in eqn. (4), $\xi(g_i,g_j)$ • 1 / ((μ/τ) + (1/μ) – 1) = $\mu\tau$ / (μ^2 + τ – $\mu\tau$). Thus eqn. (4) follows. □

3 Generalization of *BiAS* for Real-Valued Attributes

We now waive the restriction to Boolean attributes and assume $g_i = [g_{i1}, g_{i2}, \ldots, g_{id}]$, $g_{ik} \in \Re$, the set of real. With that, $c(g_i)$ is redefined to square of the L_2-norm of g_i:

$$c(g_i) = \|g_i\|^2 = \sum_{k=1}^{d} g_{ik}^2 \tag{3}$$

The frequency of joint occurrence $c(g_i \cap g_j)$ is formulated by their dot product:

$$c(g_i \cap g_j) = g_i \bullet g_j = \sum_{k=1}^{d} g_{ik} * g_{jk} \tag{4}$$

With eqn. (3) and (4), computation of $p(g_j/g_i)$, $p(g_i/g_j)$ is immediate. Also, $J(g_i,g_j) = c(g_i \cap g_j) / c(g_i \cup g_j) = c(g_i \cap g_j) / c(g_i) + c(g_j) - c(g_i \cap g_j)$. Thus, the Jaccard index (more specifically the Tanimoto coefficient [7]) and *BiAS* can be reformulated as:

$$J(g_i, g_j) = \frac{g_i \bullet g_j}{\|g_i\|^2 + \|g_j\|^2 - g_i \bullet g_j} \tag{5}$$

$$\beta(g_i, g_j) = p(g_j / g_i) * p(g_i / g_j) = \frac{[g_i \bullet g_j]^2}{\|g_i\|^2 * \|g_j\|^2} = \frac{(\sum_{k=1}^{d} g_{ik} * g_{jk})^2}{\sum_{k=1}^{d} g_{ik}^2 * \sum_{k=1}^{d} g_{jk}^2} \tag{6}$$

The physical significance of probability of a real-valued attribute may not be as straight-forward as it is for Boolean attributes. So, eqns. (5) and (6) are represented with the frequency of occurrences of an event which holds perfect without any loss of generality. Under the new formulations, $p(g_j/g_i)$, $p(g_i/g_j)$ can be more than 1 or even negative (signifying opposite similarity). Nevertheless, the absolute value of their product remains within [0,1]. Lemma (2) proves it.

Lemma 2. For any two *d*-dimensional real-valued vectors g_i and g_j,

$$0 \leq \beta(g_i, g_j) \leq 1. \tag{7}$$

Proof. Rewriting eqn. (6),

$\beta(g_i, g_j) = \frac{[g_i \bullet g_j]^2}{\|g_i\|^2 * \|g_j\|^2} = \left[\frac{g_i}{\|g_i\|} \bullet \frac{g_j}{\|g_j\|}\right]^2$. So $\beta(g_i,g_j)$ is the square of the scalar product of two unit vectors. Hence, $\beta(g_i,g_j) \geq 0$ and $\beta(g_i,g_j) \leq 1$. □

4 Connecting Lower Bounds of Different Metrics

Sometimes, a particular metric can be suitable for certain applications but the choice of threshold may not be apparent due to the physical significance not being straightforward in context of those applications. Connecting the lower bounds of different similarity metrics, particularly with different physical interpretations like similarity or mutual dependence, may help in this regard. Note that we assume $p(g_j/g_i), p(g_i/g_j) \geq \mu$, unless otherwise specified.

Relation between the lower bounds of $J(g_i,g_j)$ and $\beta(g_i,g_j)$ is derived in eqn. (2) which involves the threshold μ as well. For example, with $\mu = 0.5$ and $\tau = 0.37$ (the average of μ and μ^2), lower bound of Jaccard similarity is (0.5*0.37) / (0.5^2 + 0.37 − 0.5*0.37) = 0.185 / 0.435 = 0.425. In case τ is equal to its lower bound $(0.5)^2 = 0.25$, lower bound of $J(g_i,g_j)$ is (0.5*0.25) / (0.5^2 + 0.25 − 0.5*0.25) = 0.125 / 0.375 = 0.33. The choice of τ helps to raise the threshold of J from 0.33 to 0.425 and thus pruning additional pairs with similarities in between those two values.

From eqn. (6),

$\beta(g_i, g_j) = \frac{[g_i \bullet g_j]^2}{\|g_i\|^2 * \|g_j\|^2} = \cos^2(g_i, g_j)$. So, $cos(g_i,g_j) \geq \beta(g_i,g_j)$ as $0 \leq \beta(g_i,g_j) \leq 1$.

This establishes that mutually associative events are also similar.

Finally, we deduce the relation between lower bounds of the $J(g_i,g_j)$ and $cos(g_i,g_j)$. As $\beta(g_i,g_j) \geq \mu$, $|cos(g_i,g_j)| \geq \sqrt{\tau}$ as cosine similarity can be negative as well. Let us denote the lower bound of $|cos(g_i,g_j)|$ as $lb(|cos(g_i,g_j)|)$ and assume that $\tau = \mu^2$, where $p(g_j/g_i), p(g_i/g_j) \geq \mu$. Replacing this in eqn. (2),

$$J(g_i, g_j) \geq \frac{\mu\tau}{\mu^2 + \tau - \mu\tau} = \frac{\tau^{3/2}}{2\tau - \tau^{3/2}} = \frac{\tau^{3/2}}{\tau(2 - \sqrt{\tau})} = \frac{\sqrt{\tau}}{(2 - \sqrt{\tau})} = \frac{lb(|\cos(g_i, g_j)|)}{2 - lb(|\cos(g_i, g_j)|)}.$$

So, given the cosine similarity of g_i and g_j we can get the lower bound of $J(g_i,g_j)$.

5 Clustering Using *BiAS*

Clustering is one of the most popular and well-established unsupervised data mining techniques that deal with finding a structure in a collection of unlabeled data and determining the intrinsic grouping. The *k*-means clustering algorithm [6] is perhaps the most popular iterative solution to group events in a pre-specified *k* number of clusters. The idea is to gather all those events which lie within the preset similarity or distance from the cluster center. Thus, to ensure that *BiAS* can successfully identify a set of mutually associative events, we must prove the following two properties:

- If two events g_i and g_j happen to be mutually associative individually with a third event g_c, g_i and g_j are also mutually associative with respect to certain threshold
- If there exist pair-wise mutual association between any two events of g_i, g_j and g_c, *Jaccard* index of g_i, g_j and g_c together has a lower bound.

Lemma 3 and 4 state the lower bounds, the proofs are beyond the scope of this paper.

Lemma 3. Let g_i, g_j and g_c are three events such that $p(g_i/g_c) = \mu_1$, $p(g_c/g_i) = \mu_2$, $p(g_j/g_c) = \mu_3$ and $p(g_c/g_j) = \mu_4$. Then,

$$p(g_j / g_i) \geq \mu_2.(\mu_1 + \mu_3 - 1)/\mu_1 \tag{8}$$

$$p(g_i / g_j) \geq \mu_4.(\mu_1 + \mu_3 - 1)/\mu_3 \tag{9}$$

For example, if $\mu = 0.85$ and μ_1, μ_2, μ_3, $\mu_4 \geq \mu$, $p(g_j/g_i)$ and $p(g_i/g_j)$ will be greater than $\mu.(\mu + \mu - 1) / \mu = \mu\,(2\mu - 1) = 0.85*0.7 = 0.6$. So, if the events are clustered based on their mutual associativity with the cluster center, any pair of events within the cluster has a lower bound on their *BiAS* metric and hence the *J*. If so, lemma 4 states another lower bound for $J(g_i,g_j,g_c)$.

Lemma 4. Let g_i, g_j and g_c be three events and all of $J(g_i, g_c)$, $J(g_j, g_c)$ and $J(g_i, g_j)$ are greater than some threshold q. Then,

$$J(g_i, g_j, g_c) \geq 1 - 3(1-q)(2-q^2)/2. \tag{10}$$

E.g. with $q = 0.8$, $J(g_i,g_j,g_c) \geq 1 - 3*(1-0.8)*(2-0.64)/2 = 1 - 0.41 = 0.59$.

Thus, if a clustering algorithm ensures that an incoming event is mutually associative with the center, it is guaranteed to have mutual associativity as well as similarity with all other existing events which, in turn, maintain the overall cluster quality.

6 Conclusion

The *Bidirectional Association Similarity* ensures that similarity between two events results from true inter-dependence which we model through mutual conditional probabilities. Conventional similarity metrics, in general, just quantize the likeness between two expression vectors. It is not designed to capture similar expressions owing to mutual association between a pair of events. Apparently, *BiAS* is a stricter criterion ensuring both mutual association as well as similarity. This paper builds strong foundation of the measure, bridges connection with other well-known similarity metrics and theoretically proves the effectiveness in knowledge discovery.

We are currently working on a few promising application domains where mutual associativity is much apparent in natural phenomena and hence *BiAS* can be instrumental in knowledge discovery. Feature selection is definitely one of our interests where few other literatures [2] [3] have envisaged the effectiveness of mutual association. Also, identifying self-regulatory systems in genetics through feedback loop is another potential area to invest on.

References

1. Agrawal, R., Imielinski, T., Swamy, A.: Mining Association Rules between Sets of Items in Large Databases. In: Proc. ACM Int'l Conf. on Management of Data (SIGMOD), Washington, DC, pp. 207–216 (May 1993)
2. Blake, C., Pratt, W.: Better rules, fewer features: a semantic approach to selecting features from text. In: Proc. IEEE Intl. Conf. on Data Mining (ICDM 2001), pp. 59–66 (2001)
3. Dai, X., Jia, J., Ghaoui, L.E., Yu, B.: SBA-term: Sparse Bilingual Association for Terms. In: Fifth Intl. Conf. on Semantic Computing (ICSC 2011), pp. 189–192 (2011)
4. Everitt, B.: Cluster analysis, 3rd edn. Edward Arnold, London (1993)
5. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Englewood Cliffs, N.J (1998)
6. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1965)
7. Tanimoto, T.T.: IBM Internal Report November 17 (1957)

Inter-domain Cluster Mapping and GMCV Based Transformation for Domain Adaptation

Suranjana Samanta and Sukhendu Das

V.P. Lab, Dept. of CSE, IIT Madras, India
ssamanta@cse.iitm.ac.in, sdas@iitm.ac.in

Abstract. This paper describes an algorithm for a direct solution of domain adaptation (DA) to transform data in source domain to match the distribution in the target domain. This is achieved by formulating a transformation matrix based on the Geometric Mean of Co-Variances (GMCV), estimated from the covariance matrices of the data from both the domains. As a pre-processing step, we propose an iterative framework for clustering over data from both the domains, to produce an inter-domain mapping function of clusters. A closed form solution for direct DA is obtained from the GMCV formulation. Experimental results on real world datasets confirms the importance of clustering prior to transformation using GMCV for better classification accuracy. Results show the superior result of the proposed method of DA, when compared with a few state of the art methods.

1 Introduction

The basic assumption of the standard techniques of classification is that the training and the testing samples are drawn from the same distribution. This assumption may not always be true, specially for large-scale, real-world datasets, where we do not have clear idea about the distribution of the datasets. In certain situations, we have a very few number of training samples available from the domain of test samples for a classification task, but a large number of labeled samples are available from an auxiliary domain. During training, a very few number of training samples generally causes an over-fitting for any learning model, leading to inferior performance of classification.

The task of classification is done on test samples obtained from target domain, which often provides a few labeled/unlabeled training samples for training. On the other hand, the domain from which a large number of labeled training samples with a different distribution are available is termed as the source domain. Domain adaptation (DA) is the process of using training samples available from source domain to aid a statistical learning task to be done on testing samples obtained from target domain. There are mainly two types of DA techniques available in the literature - (a) supervised - where we have a very few number of training samples from the target domain and (b) unsupervised - where we have unlabeled training samples from target domain.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 74–81, 2013.

There has been a lot of work in DA in recent past. One solution to this problem is to weigh each instance in the source domain appropriately such that, the weighted instances of the source domain are used for training to minimize the expected loss [1], [2]. In some works, instances from source and target domains are projected onto one or more intermediate domain(s), such that the difference in distribution in two domains are less in projected domain(s). Jiang et. al. [3] and Yang et. al. [4] have proposed methods of modifying the SVM trained on samples available from source domain by introducing a bias term between source and target domains during optimization in training. There has been some work on clustering using transfer learning [5], [6]. In [5], clustering is done using KL-Divergence in a common sub-space of features while in [6], a divergence measure based on inter-cluster and intra-cluster distances has been used. The number of clusters formed in both the domains are same and known apriori.

In this paper, we propose a direct method of supervised DA, where the instances from source domain is transformed to match the distribution of the target domain using a transformation matrix. Our work uses the concept of inter-domain clustering for successful transformation of data from one domain to another. The number of clusters are automatically determined during an iterative clustering framework. We use the geometric mean of covariance (GMCV) matrices of data, for direct closed-form solution of a transformation matrix exploiting the underlying distribution in both the domains. Results are shown on datasets obtained from UCI repository [7] and a remote sensing dataset [8].

The rest of the paper is organized as follows. Section 2 gives the description of the proposed method of clustering and domain transformation. Section 3 presents and discusses the performance of the proposed methodology on real-world datasets. Section 4 concludes the paper.

2 Proposed Solution of Domain Adaptation

This paper discusses a method of DA, where instances from the source domain are transformed to match the distribution of the target domain. The proposed method, consists of two stages - (a) clustering data in both the domains and (b) Calculating Transformation Matrix using GMCV - to transform a source domain data into target domain. We assume that a minimum of one training sample per class is available from both source and target domains.

Let $\mathbf{X}, \mathbf{Y} \in \Re^{D \times n_s}$ denote the source and the target data having n_s and n_t number of samples respectively. Let $\mathbf{X}_i$ and $\mathbf{Y}_j$ denote the i^{th} and j^{th} clusters of $\mathbf{X}$ and $\mathbf{Y}$ and let x_i and y_j be the i^{th} and j^{th} instance of $\mathbf{X}$ and $\mathbf{Y}$ respectively. The entire process has been explained in the following sub-sections.

2.1 Inter-domain Clustering and Mapping

An iterative process of clustering is performed in both the domains simultaneously, where the clusters formed in one domain is influenced by the clusters formed in another domain. The proposed clustering method can be seen as an

extension of K-means clustering, where we cluster data in both domain simultaneously, where the clusters in the source domain helps to form appropriate clusters in target domain. At first, data is normalized in both the domains and $\mathbf{X}$ is then approximately aligned with the distribution of $\mathbf{Y}$ to obtain $\hat{\mathbf{X}}$, using: $\hat{\mathbf{X}} = (E_t^{-1}E_s)\mathbf{X}$, where E_s and E_t are matrices formed using the eigen-vectors of datasets $\mathbf{X}$ and $\mathbf{Y}$ respectively. $E_t^{-1}E_s$ is the higher dimensional rotation matrix which aligns the basis vectors (based on eigen-analysis) of $\mathbf{X}$ with that of $\mathbf{Y}$, such that the scatter of $\hat{\mathbf{X}}$ and $\mathbf{Y}$ are qualitatively similar to each other with significant overlap along each dimension of E_t. Let $\hat{x}_i$ be the i^{th} instance of $\hat{\mathbf{X}}$ and $\hat{\mathbf{X}}_i$ be the i^{th} cluster of $\hat{\mathbf{X}}$. Next, clusters are initialized in $\hat{\mathbf{X}}$ using K-Means clustering. The number of clusters for initialization must be considered high and it reduces in the successive iterations to give an appropriate value. Let at any iteration, K_s and K_t be the number of clusters in $\hat{\mathbf{X}}$ and $\mathbf{Y}$ respectively. Let, $\omega_s : \{1, \ldots, K_s\}$ and $\omega_t : \{1, \ldots, K_t\}$ be the sets of cluster-labels in $\hat{\mathbf{X}}$ and $\mathbf{Y}$ respectively. Let μ_s^i and μ_t^j be the mean of $\hat{\mathbf{X}}_i$ and $\mathbf{Y}_j$, $\forall i \in \omega_s$ and $\forall j \in \omega_t$.

The clusters are formed in $\mathbf{Y}$ based on the distance of each instance y_j from the means formed in $\hat{\mathbf{X}}$. The distribution of clusters are ignored in this case as it was observed to produce inferior results for classification using DA. A mapping function $\Phi : \omega_t \rightarrow \omega_s$ is calculated as,

$$\Phi(j) = i \text{ if } i = \arg\min_k dist(\mu_t^j, \mu_s^k), \forall k \in \omega_s \tag{1}$$

Here $dist(.,.)$ represents the Euclidean distance between two instances. The distribution of clusters are ignored in this case as it was observed to produce inferior results for classification in DA. Once the clusters in $\mathbf{Y}$ are formed, the clusters in $\hat{\mathbf{X}}$ are reformed based on the Bregman divergence [9] from the means of clusters of $\mathbf{Y}$. We consider the Bregman divergence of the convex function $x^T Ax$, which is given by $(x-y)^T A(x-y)$. Here, A is a positive definite matrix. For our experimentation, we consider A as $E_t^{-1}E_s$. To ensure the formation of compact clusters in $\hat{\mathbf{X}}$, inter-cluster distance is also considered. The iteration converges when the change in the cluster-labels of instances in $\hat{\mathbf{X}}$ and $\mathbf{Y}$ in successive iteration is less than 2%. The algorithm of cross-domain clustering and mapping is given in Algo. 1.

The iterative process (steps 3-9) in Algo. 1 helps to reduce the number of clusters formed in both the domains. The cluster in $\hat{\mathbf{X}}$, whose mean is not close to any of the instances in $\mathbf{Y}$ will remain outside the range of Φ (steps 4 & 5). The number of clusters formed in $\hat{\mathbf{X}}$ in step 8, will then be less than or equal to the cardinality of the range of Φ. This condition reduces the K_s which in turn will reduce K_t in the next iteration (step 4). While re-clustering $\hat{\mathbf{X}}$ (step 8), α_2 ensures that an instance $\hat{x}_i$ is assigned to a cluster with minimum Bregman divergence [9] from a cluster in $\mathbf{Y}$. On the other hand, α_1 ensures that the cluster assigned to $\hat{x}_i$ yields a compact cluster. Hence, clusters formed in both the domains will be compact due to the distance criterion (steps 4 & 6) used.

Algorithm 1. Cross-Domain Clustering and Mapping.
INPUT: Source domain: $\mathbf{X}$ and target domain: $\mathbf{Y}$.
OUTPUT: Cluster-label of instances in $\mathbf{X}$ and $\mathbf{Y}$, i.e., $\delta(x_i)$ and $\delta(y_j) \forall i = 1, 2, \ldots, n_s$ and $\forall j = 1, 2, \ldots, n_t$ and the mapping function $\Phi : \omega_t \rightarrow \omega_s$.

1: Normalize $\mathbf{X}$ and $\mathbf{Y}$ and align source domain to obtain $\hat{\mathbf{X}} = (E_t^{-1} E_s)\mathbf{X}$.
2: Cluster $\hat{\mathbf{X}}$ with K_s (large) number of clusters using K-means, forming means of clusters μ_s^i for $i \in \omega_s$.
3: **repeat**
4: Cluster instances in $\mathbf{Y}$ as: $\delta(y_j) = \arg\min_k dist(y_j, \mu_s^k)$ and obtain $\mu_t^j \ \forall j$.
5: Calculate the cross-domain mapping function, Φ, as: $\Phi(j) = \arg\min_k dist(\mu_t^j, \mu_s^k), \forall k \in \omega_s$ and $\forall j \in \omega_t$.
6: Calculate cluster compactness of $\hat{\mathbf{X}}_k$ as: $\alpha_1(k) = \frac{1}{n_s^k} \sum_{\hat{x} \in \hat{\mathbf{X}}_k} dist(\hat{x}, \mu_s^k), \forall k \in \omega_s$.
7: Calculate Bregman divergence [9] of each instance of $\hat{\mathbf{X}}$ with respect to each of the clusters formed in $\mathbf{Y}$, as: $\alpha_2(i,k) = (\hat{x}_i - \mu_t^k)^T (E_t^{-1} E_s)(\hat{x}_i - \mu_t^k), \forall i = 1, 2, ..., n_s$ and $k \in \omega_t$.
8: Re-cluster $\hat{\mathbf{X}}$, after normalizing α_1 and α_2, as: $\delta(\hat{x}_i) = \arg\min_k [\lambda \alpha_2(i,k) + (1-\lambda)\alpha_1(\Phi(k))], \forall k \in \omega_t$ and $\forall i$. ($\lambda = 0.6$, found empirically).
9: Recompute means of all clusters: $\mu_s^k, \forall k \in \omega_s$.
10: **until** Convergence (i.e., $\delta(\hat{x}_i)$ & $\delta(y_j)$ are unchanged)

2.2 Estimation of Transformation Matrix Using GMCV

We transform data from $\mathbf{X}$ considering each cluster at a time. We consider a transformation, which ensures that clusters formed in $\hat{\mathbf{X}}$ are same as that in $\hat{\mathbf{X}}$, i.e., $\delta(x_i) = \delta(\hat{x}_i), \forall x_i \in \mathbf{X}$. If $\Phi(j) = i$, then $\mathbf{X}_i$ is transformed to match the distribution of $\mathbf{Y}_j$. Each of the clusters formed in $\mathbf{X}$ and $\mathbf{Y}$ are hyper-ellipsoidal in nature, which can be considered as clusters following a Gaussian distribution. It becomes less erroneous to calculate the transformation matrix if the given source and the target distributions can be modeled by a similar distribution but with different parameters. Let, $\mathbf{X}_i$ be transformed using a transformation matrix W_i to match the distribution of $\mathbf{Y}_j$ ($\Phi(j) = i$). We use geometric mean of covariance matrices of data from $\mathbf{X}$ and $\mathbf{Y}$ to calculate W_i. Let, $\tilde{\mathbf{X}}_i$ be the transformed source domain cluster: $\tilde{\mathbf{X}}_i = W_i \mathbf{X}_i$.

Let C_s^i and C_t^j be the covariance matrices of $\mathbf{X}_i$ and $\mathbf{Y}_j$ respectively. The covariance matrix of $\tilde{\mathbf{X}}_i$ can be represented as $W_i C_s^i W_i^T$, which should be equal to C_t^j as the distribution of $\mathbf{Y}_j$ should be same as $\tilde{\mathbf{X}}_i$. This yields a quadratic expression of W_i: $W_i C_s^i W_i^T = C_t^j$. If W_i is symmetric, we get $W_i C_s^i W_i = C_t^j$ and W_i becomes the geometric mean of covariances (GMCV): $(C_s^i)^{-1}$ and C_t^j [10]. Let $C_s^i = U^T U$, where U is an upper triangular matrix obtained by performing Cholesky's decomposition of C_s^i. Let, $V = U^{-1}$ and $R^2 = U C_t^j U^T$, i.e., R is the

matrix square-root of $UC_t^j U^T$. Then, the solution to W_i based on GMCV, is obtained as the following expression (substitution yields the proof ⊠):

$$W_i = VRV^T \tag{2}$$

Eqn. 2 gives a direct solution to DA, to obtain $\tilde{\mathbf{X}}$ based on GMCV.

2.3 Result on a Toy Dataset for Illustration

To explain the steps of the proposed algorithm, we consider a simple example of data distribution in source ($\mathbf{X}$) and target ($\mathbf{Y}$) domains in $\Re^2$. In Fig. 1 (a) the cyan and the green points denote the instances from $\mathbf{X}$ and $\mathbf{Y}$. Fig. 1 (b) shows the overlap of $\hat{\mathbf{X}}$ with $\mathbf{Y}$, where the approximate alignment is performed using $E_t^{-1}E_s$ (step 1 in Algo. 1). The result of cross-domain clustering is shown in Fig. 1 (c). The yellow, light blue and orange points denote the data in three clusters formed in $\mathbf{X}$, while blue, magenta and black points denote the data in three clusters formed in $\mathbf{Y}$. Fig. 1 (d) shows the transformed source domain data, $\tilde{\mathbf{X}}$, in red points obtained using the proposed method (Eqn. 2). Fig. 1 (e) shows the transformed source domain data when clustering is avoided as the pre-processing step, producing inferior transformation compared to that shown in Fig. 1 (d). Transformed source domain data marked in red and violet in Figs. 1 (d) & (e), overlap the data in target domain, marked in green points. Fig. 1 (f) gives two plots shown against the number of clusters formed in each iteration, showing the change in: (i) ratio of cluster-labels in $\hat{\mathbf{X}}$ and (ii) the average KL-Divergence between the clusters formed in $\hat{\mathbf{X}}$ and $\mathbf{Y}$, for cross-domain clustering. Both monotonically decrease with increase in iterations (Algo. 1). The KL divergence measure with respect to the target domain, for the original, aligned and transformed source domains in Figs. 1 (a), (b), (d) and (e) are 14.8191, 0.5451, 0.0250 and 0.6240 respectively. Least value of KL-Divergence is obtained for Fig. 1 (d), which shows the efficiency of the proposed method, specifically when inter-domain clustering is used.

3 Experimental Results and Performance Analysis

We evaluate the performance of the proposed method on real world datasets obtained from UCI repository [7] and landmine dataset [8]. We describe the two sets of experimentation done to exhibit the efficiency of the proposed method.

Landmine dataset has been used for DA by Shi et. al. [8]. This dataset has 5 source domains (Dataset 20 - Dataset 24) with different distributions than the target domain and one source domain (Dataset 6-10) with similar distribution with the target domain. The dataset has two classes and the feature set has nine dimensions. Further details of the dataset is given in [8]. We conduct a 30-fold experimentation and consider one instance from every class of the target domain randomly for training purpose, which is similar to the experimental setup used by Shi et al. [8]. In this case, due to the presence of only one training sample

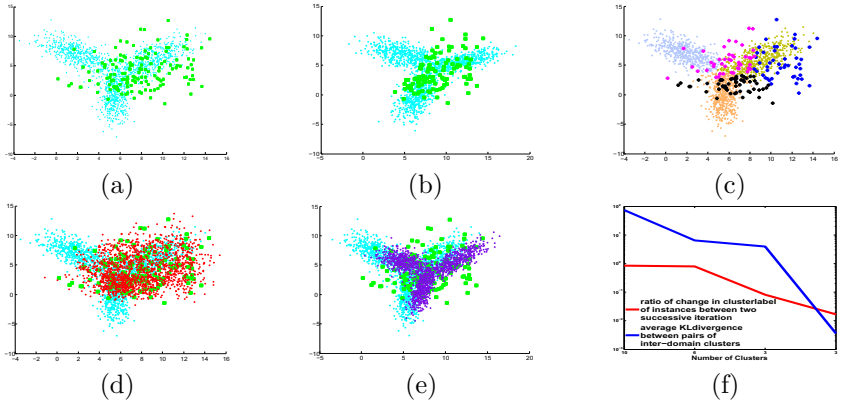


Fig. 1. Scatter plots (a)-(e) showing the effect of proposed method using a toy dataset. (a) cyan - original source domain, green - target domain; (b) cyan - aligned source domain (step 1 in Algo. 1), green - normalized target domain; (c) three clusters (shown in light blue, orange and mustard) formed in source domain and three clusters (shown in blue, magenta and black) formed in target domain using Algo. 1; (d) red - transformed source domain using the proposed method of clustering and GMCV transformation which completely overlaps with the distribution of target domain; (e) violet - transformed source domain using only GMCV transformation which partly overlaps with the distribution of the target domain; (f) Plots showing changes in average KL-Divergence between pair of inter-domain clusters with each iteration in blue and the ratio of change in cluster-labels in successive iterations in red, shown against the number of clusters in each iteration.

in target domain per class, application of the process of iterative clustering was not meaningful. Hence, we transform source data considering samples from all classes together, without using cross-domain clustering as the pre-processing step. The average accuracy of 30-fold experimentation using SVM classifier with Gaussian kernel is given in Table 1. We report the results in [8] (columns 2, 3 & 4), whereas the performance of ASVM [4] and CD-SVM [3] (column 5 & 6) are obtained using the code given in [11]. Results shows that our method (column 7) performs better than the state of the arts as reported, for all datasets having signification disparity of distribution between two domains (Dataset 20-24).

In the 2^{nd} set of experimentation, we consider 3 datasets - ionosphere, mushroom and WDBC from UCI repository [7]. We split each dataset into source and target domains using the method explained in [2]. We observe the accuracy of SVM classifier with Gaussian kernel by varying the fraction of training samples obtained from the target domain (0.3 to 0.7). Comparative studies are done using ASVM [4], CD-SVM [3] and KLIEP [1]. Two methods of experiments are performed using GMCV based transformation: (i) C-GMCV - with inter-domain clustering & (ii) GMCV - without inter-domain clustering. Results in Fig. 2 shows the classification accuracy using different techniques of DA for varying number of training samples obtained from the target domain. The red and the green curves show the classification accuracy using the proposed

Table 1. Classification accuracy (in %-age) of Landmine dataset [8] using different techniques of domain adaptation. Best classification accuracy is highlighted in bold.

Source Data	Source only [8]	TrAdaboost [2], [8]	AcTraK [8]	ASVM [4]	CD-SVM [3]	Proposed method
Dataset 20	57	89.76	94.49	94.31	90.77	**94.58**
Dataset 21	57	86.04	94.48	94.18	94.02	**94.58**
Dataset 22	57	90.50	94.49	94.44	94.56	**94.60**
Dataset 23	57	88.42	94.49	94.08	83.97	**94.58**
Dataset 24	57	90.70	94.49	94.34	93.57	**94.58**
Dataset 6-10	57	**94.76**	94.70	94.55	94.58	94.66

methods: C-GMCV and GMCV respectively. When the training samples from both the domains are taken together for training, we term it as the naive combination, whose performance is shown by the black curve. The blue curves show the performances when training samples from only one of the domains ($\mathbf{X}$ or $\mathbf{Y}$) is used for training the classifier. The cyan, brown and magenta curves show the classification accuracy using different methods published for DA: KLIEP [1], ASVM [4] and CDSVM [3]. Plots for three datasets show that the proposed method of C-GMCV gives the best result. The results also show that the inclusion of proposed method of inter-domain clustering enhances the classification performance of DA.

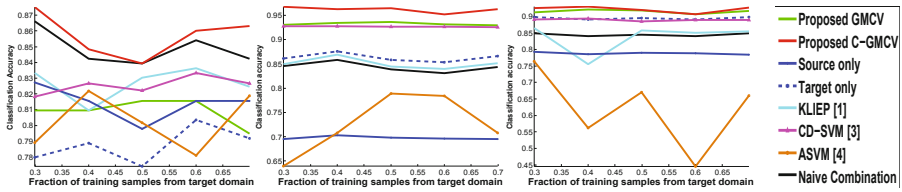


Fig. 2. Classification accuracy of (a) Ionosphere, (b) Mushroom and (c) Wdbc using different techniques of DA with varying fraction of training samples obtained from target domain. (d) shows the color indices indicating the methods used to get different performance curves. The proposed C-GMCV (in red curve) gives the best accuracy.

4 Conclusion

We propose an algorithm for a direct solution of domain adaptation based on geometric mean of covariance matrices (GMCV) of data in both the domains, which exploits the cluster information present in the data. Results demonstrated in the paper show that the proposed method is better than that reported in published literature in almost all the cases. The work can be extended for better object categorization and face recognition.

References

1. Sugiyama, M., Nakajima, S., Kashima, H., von Bünau, P., Kawanabe, M.: Direct importance estimation with model selection and its application to covariate shift adaptation. In: Neural Information Processing Systems, pp. 1962–1965 (2007)
2. Dai, W., Yang, Q., Xue, G.R., Yu, Y.: Boosting for transfer learning. In: International Conference on Machine Learning, pp. 193–200 (2007)
3. Jiang, W., Zavesky, E., Fu Chang, S., Loui, A.: Cross-domain learning methods for high-level visual concept classification. In: International Conference on Image Processing, pp. 161–164 (2008)
4. Yang, J., Yan, R., Hauptmann, A.G.: Cross-domain video concept detection using adaptive SVMs. In: International Conference on Multimedia, pp. 188–197 (2007)
5. Dai, W., Yang, Q., Xue, G.R., Yu, Y.: Self-taught clustering. In: International Conference on Machine Learning, pp. 200–207 (2008)
6. Bhattacharya, I., Godbole, S., Joshi, S., Verma, A.: Cross-guided clustering: Transfer of relevant supervision across domains for improved clustering. In: International Conference on Data Mining, pp. 41–50 (2009)
7. Asuncion, A., Newman, D.H.: UCI machine learning repository (2007)
8. Shi, X., Fan, W., Ren, J.: Actively transfer domain knowledge. In: Daelemans, W., Goethals, B., Morik, K. (eds.) ECML PKDD 2008, Part II. LNCS (LNAI), vol. 5212, pp. 342–357. Springer, Heidelberg (2008)
9. Banerjee, A., Merugu, S., Dhillon, I.S., Ghosh, J.: Clustering with Bregman Divergences. Journal of Machine Learning Research 6, 1705–1749 (2005)
10. Lawson, J.D., Lim, Y.: The geometric mean, matrices, metrics, and more. The American Mathematical Monthly 108(9), 797–812 (2001)
11. Duan, L., Xu, D., Tsang, I.W.H.: Domain adaptation from multiple sources: A domain-dependent regularization approach. IEEE Transaction Neural Network Learning System 23(3) (2012), `http://vc.sce.ntu.edu.sg/transfer-learning-domain-adaptation/domain-adaptation-home.html`

Structural Feature Based Classification of Printed Gujarati Characters

Mukesh Goswami[1] and Suman Mitra[2]

[1] Department of IT, Faculty of Technology, D.D. University
mgoswami.it@ddu.ac.in
[2] Dhirubhai Ambani Institute of Information and Communication Technology
suman_mitra@daiict.ac.in

Abstract. This paper presents a Structural feature based method for classification of printed Gujarati characters. The ability to provide incremental definition of characters in terms of its native components makes the proposal unique and versatile. It deals with varied sizes, font styles, and stoke widths. The features are validated on subset of machine printed Gujarati characters using a simple rule based classifier and the initial results are encouraging.

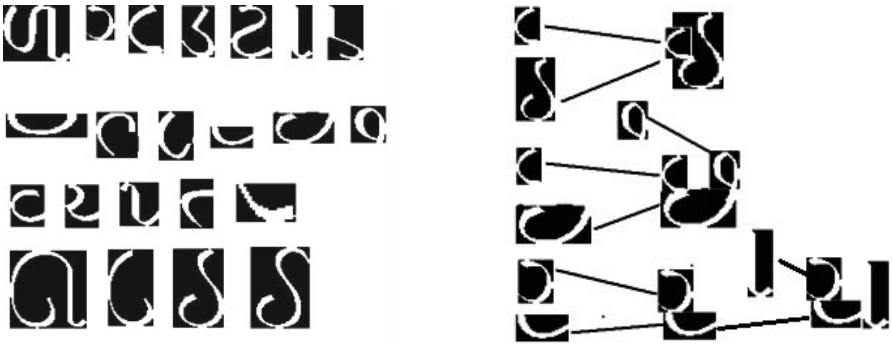
Keywords: Structural Features, Character Classification, Gujarati Characters.

1 Introduction

Development of Optical Character Recognition (OCR) technology for Indian text is more challenging then western text because of the complex character set as well as existence of modifiers and joint characters. Motivated by this many researchers have started working on the OCR for Indian text over a decade ago and it is still in moderate phase for many languages. The majority of the work found in the literature for classification of characters from Indian text can be divided into two major streams 1) Structural feature based approach and 2) Transform domain feature based approach. For many languages like Hindi, Bangla, Gurumukhi, Oriya etc., the structural features along with simple rule based classifier like decision trees have performed well[1–4]. On the other hand, the south Indian scripts like Tamil, Telugu, and Kannada, where it is difficult to identify characters from the general shape and structure, the transform domain features along with sophisticated classifiers like Neural Networks, Support Vector Machine etc. have done well[5–8]. Some work on classification of Gujarati text using transform domain features can be found in the literature[9–12]. Gujarati is derived from the ancient Devanagari script and having close resemblance with other north Indian script, primarily Hindi. The major difference between Gujarati and other north Indian script is the absence of "Shirorekha", a head line running through all the characters forming the word. Even though Gujarati characters have well define shape and structure, but no structural features based method considered so far for the classification of Gujarati text. This paper attempts to propose a

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 82–87, 2013.

structural feature extraction method and its use in classification of subset of Gujarati characters. The rest of the paper is organized as follow Section 2 describes the brief of various steps used in structural feature extraction method and its use in classification, followed by simulation result and conclusion in Section 3 and 4 respectively.



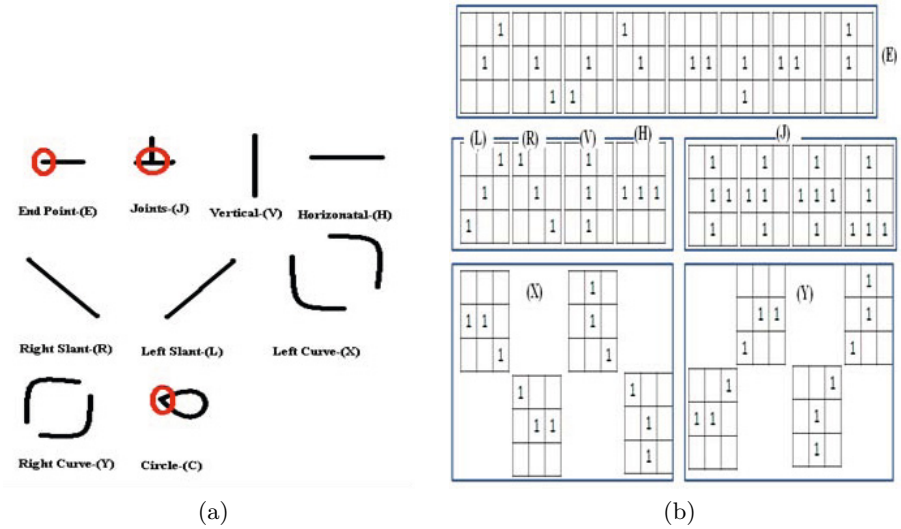
(a) (b)

Fig. 1. (a) Subset of components formulating majority of Gujarati character set (b) Formation of characters from the components

2 Proposed Methodology

The Fundamental idea for the formation of Gujarati characters is to consider every characters as combination of some components (also referred as strokes in some literature) connected in specific ordered. This break up of characters in components helps in exploiting the reusable components in similar looking characters. A total of 30 components are identified which formulates majority of characters from Gujarati character set. Some of these components are shown in Fig. 1a. An example of how characters are formulated by combining the components in a specific order is shown in Fig. 1b. Every component in turn can be defined as some sequence of primitive strokes. Set of primitive strokes that formulate a complex component is shown in Fig. 2a. All primitive strokes are represented by string symbol (as shown in Fig. 2a). Thus the component is described by sequence of string symbols representing primitive strokes and the character in turn can be defined as the sequence of components occurring in some specific order (as shown in Fig. 3). A complete system for identification of native components and formation of characters as ordered set of native components is described as follows.

Preprocessing and Applying 3x3 Pattern Mask. Input to the system is a binary character image which is passed through various preprocessing stages like noise removal, resizing and thinning. Thinning is one of the most important preprocessing steps that converts elongated character image into one pixel wide thinned image that preserves the original shape of the character without



(a) (b)

Fig. 2. (a) Set of primitive components (b) Set of 3x3 pattern mask to identify primitive components

losing the connectivity. Also it facilitates more elegant description of component shapes.Various 3x3 pattern masks (as shown in Fig. 2b) are applied to detect primitive strokes from the thinned image. Every detected primitive stroke is replaced by predefined numerical code. Thus, an MxN input image is converted into an MxN matrix of numerical codes.

Scanning, Component Separation and Symbol Array Generation. In order to preserve the component ordering, MxN matrix of numerical codes is scanned. Since Gujarati writing is left to right and top to bottom, zigzag scanning is used to find the start point. Modified contour tracing algorithm that selects the next component in clockwise direction is designed to scan components in nearly writing flow order starting from the start point. For example writing flow order of components present in the character image shown in Fig. 3 is $\{C_1C_2C_3C_4C_5\}$. During scanning, numerical code(s) are replaced with equivalent character symbols of primitive strokes to obtain symbol string representation. The components are separated from the symbol string by using end points symbol (E) and junction points symbol (J) as separators, thus every component is represented by string of symbol as shown in Fig. 3.

Noise Removal from the Symbol String and Identification of Component. A noise removal technique is designed to further fine tune the symbol string representation. Regular Expression (RE) matching based method is designed to identify every component, represented by a symbol string. It takes component string and RE file as input. The RE file defines set of regular expression for each component class. The method then generates target component string using all RE's present in the file and gives a matching score close to 1.0

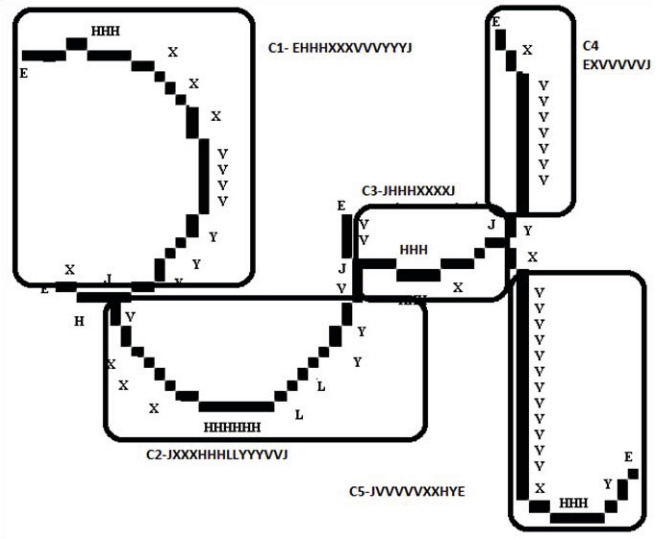


Fig. 3. Formation of Character as sequence of components described as sequence of primitive strokes represented by string symbols

if the string is fully generated by the given RE. Finally, the component represented by RE having maximum matching score is selected as target component and replaced by its id. Thus the character is finally represented by an ordered set of component ids'.

Classification. A simple rule based algorithm is designed to check the discriminating power of the proposed structural features. Every component is assigned some weight depending on how much it contributes in the recognition of a given character. Accumulated weights for all characters are maintained in the form of a vector. The algorithm proceeds as follows.

- Initialize the accumulated weight vector $W = \{w_{1,}w_{2,\ldots\ldots}w_m\}$ to zero, for all m number of character classes.
- Get Input character image and pass it through the feature extraction stages to obtain the ordered set of components ids' $C = \{c_{1,}c_{2,}\ldots c_n\}$.
- Select the appropriate rule set $R(n)$ for different character classes depending on the number of component(n). Rule R_x for given character class x is a set of component value pair $< r_{xi}, v_{xi} >$ where v_{xi} is weight of component r_{xi} with respect to character class x.

  1. For every rule R_i corresponding to class i in the selected rule set $R(n)$
  2. **For** every component value pair $< r_{ij}, v_{ij} >$ in R_i
  3 **if** $r_{ij} \in C$ **then** increase the accumulated weight w_i by v_{ij}
  4. **else** component r_{ij} doe not contribute in defining character i.
- Select the Character class y with maximum of accumulated weights (w_y) as predicted class

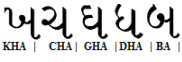
3 Simulation Results

The proposed method was tested on moderate size database of 4000 machine printed character symbols from 20 different pages of 4 different machine printed books. The result of the experiment are shown in Table 1. It is evident from the results that even though the classifier being simple it gives very high accuracy for simple characters having 1, 2 or 3 components (around 95%) which constitute almost half the character set. The accuracy drops as the number of components increases in the character. When number of components are large, it becomes difficult to assign optimum weights to all character component manually. Thus a mechanism is needed to automatically find the optimum weights to components. Also the classifier does not consider ordering of component, which is crucial and may lead to better accuracy.

4 Conclusions and Future Work

A structural feature extraction method is proposed for classification of printed Gujarati text and tested on machine printed character symbol data set of size 4000. The accuracy obtained is very high for simple character with small number of components, however method does require some improvement like finding the component weights automatically and exploiting the order of component. The salient feature of the proposal is its ability to provide incremental definition of

Table 1. Accuracy of various character class

Sr.No.	No. of Components	Character Class	Accuracy
1	1	 GA | TTA | DDA | LA | LLA	GA=97% TTA=98% DDA=97% LLA=91% LA=96%
2	2	 TTHA | DDHA |	TTHA=91.66% DDHA=81.25%
3	3	 PA | YA | RA | VA | TA |	PA=95% YA=90% RA=96% VA=88% TA=90%
4	3	 U | UU | II | I | DA	U=87% UU=66.66% II=97% I=95% DA=95%
5	4	 CHA | SSA	CHA=93% SSA=60%
6	5	 KHA | CHA | GHA | DHA | BA |	KHA=83% CHA=47% GHA=86.66% DHA=98% BA=94%
7	6	 BHA | MA |	BHA=85% MA=66%
8	7	 SA | A |	SA=86% A=66%

characters in terms of components. As the scope of the work is to evaluate the proposed feature hence comparision with other methods is omited.

References

1. Sinha, R.M.K., Mahabala, H.N.: Machine Recognition of Devanagari Script. IEEE Transactions on Systems, Man, and Cybernetics 9, 435–441 (1979)
2. Chaudhuri, B.B., Pal, U.: A complete printed Bangla OCR system. Pattern Recognition 31, 531–549 (1998)
3. Pal, U., Chaudhuri, B.B.: Printed Devanagiri Script OCR System. Vivek 10, 12–24 (1997)
4. Lehal, G.S., Singh, C.: A Gurmukhi script recognition system. In: Proc. of the 15th International Conference on Pattern Recognition (ICPR), pp. 557–560 (2000)
5. Aparna, K.G., Ramakrishnan, A.G.: A complete Tamil optical character recognition system. In: Lopresti, D.P., Hu, J., Kashi, R.S. (eds.) DAS 2002. LNCS, vol. 2423, pp. 53–57. Springer, Heidelberg (2002)
6. Jawahar, C.V., Pavan Kumar, M.N.S.S.K., Ravi Kiran, S.S.: A bilingual OCR for Hindi-Telugu documents and its applications. In: Proceedings of Seventh International Conference on Document Analysis and Recognition, pp. 408–412. IEEE Computer. Soc. (2003)
7. Manjunath, V.N., Aradhyal, P.S., Kumar, G.H., Noushathl, S.: Fisher linear discriminant analysis based technique useful for efficient character recognition. In: Proc. of the 4th International Conference on Intelligent Sensing and Information Processing, pp. 49–52 (2006)
8. Ashwin, T., Sastry, P.: A font and size-independent ocr system for kannada documents using SVM. Sadhana 27 (2002)
9. Antani, S., Agnihotri, L.: Gujarati character recognition. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition, ICDAR 1999 (Cat. No.PR00318), pp. 418–421. IEEE (1999)
10. Dholakia, J., Negi, A., Rama Mohan, S.: Progress in Gujarati Document Processing and Character Recognition. In: Govindaraju, V., Setlur, S. (eds.) Guide to OCR for Indic Scripts: Document Recognition and Retrieval, pp. 73–95. Springer Publishing Company (2009)
11. Hassan, E., Chaudhury, S., Gopal, M., Dholakia, J.: Use of MKL as symbol classifier for Gujarati character recognition. In: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems DAS 2010, pp. 255–262. ACM Press, New York (2010)
12. Goswami, M. M., Prajapati H. B., Dabhi V. K.: Classification of printed Gujarati characters using SOM based k-Nearest Neighbor Classifier. In: IEEE International Conference on Image Information Processing, ICIIP 2012. pp. 1-5. IEEE (2013)

Mining Anomalous Sub-graphs in Graph Data Using Non-negative Matrix Factorization

N.N.R. Ranga Suri[1], Musti Narasimha Murty[2], and Gopalasamy Athithan[1,3]

[1] Centre for AI and Robotics (CAIR), Bangalore, India
{rangasuri,athithan.g}@gmail.com
[2] Dept of CSA, Indian Institute of Science (IISc), Bangalore, India
mnm@csa.iisc.ernet.in
[3] Presently working at Scientific Analysis Group (SAG), Delhi, India

Abstract. Mining graph data has been an important data mining task due to its significance in network analysis and many other contemporary applications. Detecting anomalies in graph data is challenging due to the unsupervised nature of the problem and the size of the data itself to be dealt with. Recent research efforts in this direction have explored graph data for identifying anomalous nodes and anomalous edges of a given graph. However, in many real life applications where the data is inherently networked in nature, the requirement is to detect anomalous sub-graphs with distinguishing characteristics such as *near cliques*, etc. In this context, we propose a novel method for addressing the anomalous sub-graph mining problem through community detection by employing the non-negative matrix factorization technique. Anomalous sub-graphs are identified by applying some existing techniques on the detected communities for measuring their deviation from the normal characteristics. We demonstrate the effectiveness of the proposed method through experimental evaluation on various benchmark graph data sets.

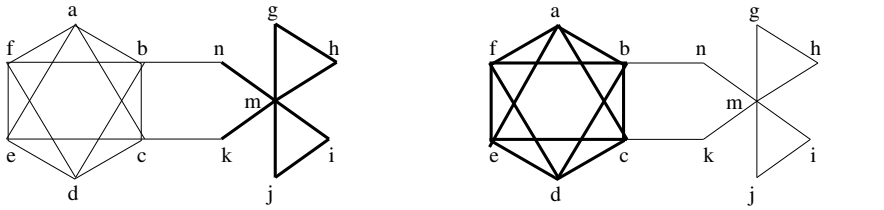
Keywords: Data mining, Mining graph data, Anomalous sub-graphs, Community detection.

1 Introduction

An emerging research problem related to graph mining is to discover anomalies, also known as outliers, in graph data [2,10,3]. The objective is to identify the sub-graphs displaying anomalous characteristics in the input graph. This problem assumes significance as finding out a close group of individuals in a social network (forming a clique like pattern) or identifying minimal interactions among a set of nodes in a communication network (forming a tree like pattern) helps in further analysis of this sub-set data in an objective manner. Thus, anomaly detection in graph data turns out to be an important mining task.

A recent work on anomaly detection in graph data [1], makes use of the notation of *egonet* of a node defined as the induced sub-graph of its 1-step neighbors as shown in Fig. 1(a). According to this work, it is found that the number of

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 88–93, 2013.



(a) Egonet of node m (bold lines) (b) An arbitrary sub-graph (bold lines)

Fig. 1. Examples of sub-graphs of different types

nodes and the number of edges of a normal egonet satisfy a power law relationship. As a result, the anomalous egonets (*near cliques* and *near stars* as defined in [1]) are identified by employing the established power law relationship.

It is clear that this result is applicable only to identify anomalous sub-graphs that are egonets structurally. However, there can exist various anomalous sub-graphs which are arbitrary sub-graphs of a given graph, not necessarily being sub-graphs of the egonet type. For example, an arbitrary sub-graph as shown in Fig. 1(b) qualifies to be an anomalous sub-graph (near clique) that is not an egonet structurally.

The above discussion clearly motivates the strong need to develop a generic approach to identify anomalous sub-graphs. In this paper, we propose a novel method for identifying anomalous arbitrary sub-graphs of a given graph. Similar to [1], the novel method is intended to detect *near clique* and *near tree* types of anomalies. In contrast to a near star anomaly, a near tree anomaly represents a rooted hierarchical structure, which is not an egonet. Firstly, the set of all possible connected sub-graphs is determined through community detection [4] by employing the Non-negative Matrix Factorization (NMF) technique [5]. The rationale behind this task is that the sub-graphs identified through community detection are considered to be the best candidates for detecting anomalies in graphs. As listing out all possible connected sub-graphs of a large graph leads to combinatorial explosion, the same is achieved through community detection. Then, each one of these sub-graphs is subjected to the power law relationship based anomaly detection procedure [1].

The following section gives a brief summary of various methods developed for anomaly detection in graph data. Section 3, describes the novel method being proposed for mining anomalous sub-graphs. Details on the experimental evaluation of the proposed method along with the results obtained are furnished in Section 4. Finally, Section 5 concludes the paper with a discussion on the proposed method along with a few directions for future work.

2 Related Work

There have been some research efforts made in the recent past for detecting various types of anomalies in graph data. The method proposed in [8] is an early one

using the Minimum Description Length (MDL) principle. The main idea of this method was that sub-graphs containing many common sub-structures are generally less anomalous than sub-graphs with few common sub-structures. Thus, it is suitable mainly for applications involving many common sub-structures such as the graphs describing the atomic structure of various chemical compounds. Similarly, the method proposed in [9] is meant for anomalous link (edge) discovery defining a novel edge significance measure. More recently, a method named *OddBall* was proposed [1] for detecting anomalous sub-graphs using the egonet notation, as depicted in Fig. 1(a). According to this method, the number of nodes N_i and the number of edges E_i of the egonet G_i of a node i follow a power law relationship, named the Egonet Density Power Law (EDPL), defined as

$$E_i \propto N_i^{\alpha}, \quad 1 \leq \alpha \leq 2. \tag{1}$$

Consequently, anomalous egonets (near cliques and near stars) are determined by measuring the amount of deviation from the power law relationship.

As discussed in the previous section, detecting communities in graph data is an important related graph mining task. According to the literature, communities in a graph represent groups of nodes with dense connectivity within each group [4]. Among the existing methods for detecting communities in graph representation of network data, the method proposed in [11] is a recent one employing the Non-negative Matrix Factorization (NMF) technique due to its powerful interpretability and close relationship with the clustering methods. The NMF algorithm [5] was basically proposed for finding factors of a matrix as $G \approx WH$, highlighting the philosophy of 'parts-based representation of objects'. Sparseness of the data is another important aspect managed effectively by the NMF technique. Thus, the same is employed here for identifying various connected sub-graphs of the input graph.

3 Proposed Method

As stated earlier, the objective is to identify various sub-graphs with anomalous characteristics in a given graph. This is achieved through first determining various candidate sub-graphs in the input graph and then putting them through an anomaly detection procedure as per the scheme shown in Fig. 2.

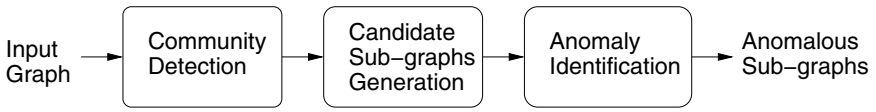


Fig. 2. Proposed scheme for detecting anomalous sub-graphs

As mentioned in Section 2, non-negative matrix factorization (NMF) [5] is a promising method for community detection [4] in graph data. Accordingly, the

adjacency matrix (G) corresponding to the input graph is subjected to community detection using the NMF procedure [5] by assigning a user specified value for the number of communities (indicated by k) to be detected. This results in a mapping W (given by $G \approx WH$) of the N nodes in G to various communities ($\{G_1, G_2, ..., G_k\}$) with varying degree of membership. Here, a membership value $w_{ij} \in W$ indicates the membership of the node n_i in G to the community G_j (j^{th} column of W). Applying a suitable threshold value (indicated by t) on the membership values (elements of W), crisp communities (sub-graphs) are determined. Let N_i be the number of nodes and E_i be the number of edges present in the sub-graph G_i. We now define the following two constraints to prune the set of crisp sub-graphs producing the candidate set of l ($0 \leq l \leq k$) sub-graphs.

(A) *Sub-graph non-triviality constraint*: Minimum number of nodes required in a non-trivial sub-graph is set as 3, i.e., $N_i \geq 3$.
(B) *Sub-graph connectivity constraint*: Minimum number of edges present in a connected sub-graph, i.e., $E_i \geq (N_i - 1)$.

For exploring the anomalous characteristics of the l candidate sub-graphs satisfying the above constraints, the power law relationship given in Equation 1 is considered here. In the context of mining anomalous arbitrary sub-graphs, this relationship is renamed as Sub-graph Density Power Law (SDPL) to reflect the property satisfied by the arbitrary sub-graphs of the input graph. Accordingly, a scatter plot with E_i versus N_i on the log-log scale is produced depicting the l sub-graphs along with the least squares fitting line. Thus, the outlierness score of a sub-graph G_i is computed as the distance to the fitting line as defined in [1].

$$out\_score(G_i) = \frac{max(E_i, CN_i^{\alpha})}{min(E_i, CN_i^{\alpha})} log(|E_i - CN_i^{\alpha}| + 1) \tag{2}$$

Finally, the anomalous sub-graphs are indicated in the scatter plot showing their deviation from the fitting line. A summary of various computational steps involved in the proposed method is furnished in Algorithm 1.

Algorithm 1. Mining anomalous sub-graphs in graph data.

Input: A graph with N nodes and E edges given by its adjacency matrix G.
Output: List of top ranked anomalous sub-graphs.

1: Factorize the adjacency matrix using the NMF procedure as $G \approx WH$.
2: Determine k communities $\{G_1, G_2, ..., G_k\}$ using W matrix.
3: Apply the membership threshold t on W to generate crisp communities.
4: Apply the sub-graph constraints A and B to determine the candidate set.
5: Produce the E_i versus N_i scatter plot along with least squares fitting line.
6: Measure the deviation of each candidate sub-graph G_i using Equation 2.
7: Obtain a ranked sequence of the sub-graphs as per their outlier scores.

4 Experimental Evaluation

For the purpose of evaluating the proposed method, we have considered certain real life graph data sets from the SNAP repository [6]. As the basic task here is

to detect the communities determined by the connectivity structure, the graphs employed in this experimentation are undirected and without any weights on the edges, as per the details furnished in Table 1.

Table 1. Details of the experimentation with the SNAP graph data sets

Input Details			Parameter Values		Observations	
Data Set Name	# Nodes (N)	# Edges (E)	# Comm. (k)	Membership threshold (t)	# Candidate sub-graphs (l)	Highest score
Facebook-107	1,034	53,498	100	0.01	100	4.1549
AS20000102	6,474	13,233	500	0.001	50	2.6012

The initial experiments were carried out on the Facebook graph [7] corresponding to the node '107', the largest node graph. This graph was subjected to the proposed sub-graph anomaly detection method resulting in the scatter plot as shown in Fig. 3(a), as in [1]. This plot represents the anomalous sub-graphs in the form of triangle points, with the area of each triangle point representing the amount of deviation of the corresponding sub-graph, as measured using Equation 2. Triangle points below the fitting curve (solid line) indicate near tree anomalies, while the ones above the fitting curve indicate near clique anomalies.

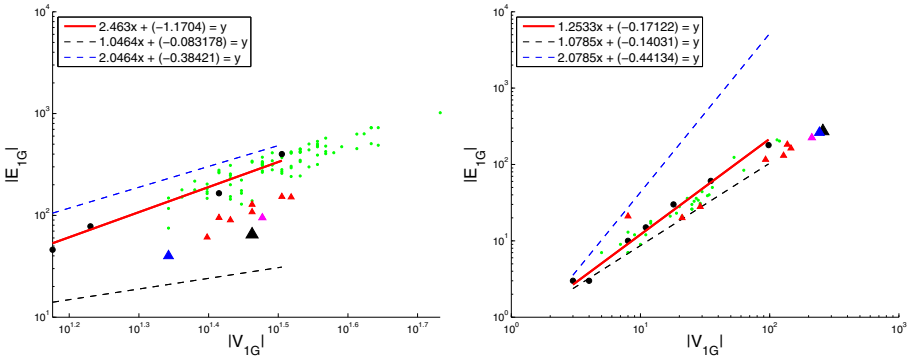
Similar experimentation was carried out on the 'AS20000102' graph from the Autonomous Systems collection [6] producing the results as shown in Fig. 3(b).

Table 1 provides the details on the values set for the parameters of the proposed algorithm (k and t) in this experimentation. It also indicates the number of candidate sub-graphs surviving after applying the proposed two constraints on the crisp communities resulted through the threshold application on the community membership values. Also shown is the highest anomaly score obtained corresponding to each graph data set. Referring to the input graph structure, it was found that many of the anomalous sub-graphs detected in this process are not egonets structurally, demonstrating the merit of the proposed method.

5 Conclusion and Future work

A novel method for mining anomalous sub-graphs in graph data has been proposed here through community detection by employing the NMF technique. The proposed method identifies anomalous sub-graphs by subjecting the detected communities to the power law constraints defined in the OddBall method. While the OddBall method can detect anomalous sub-graphs that are egonets structurally, the novel method can detect any arbitrary sub-graph of the input graph having anomalous characteristics. Thus, the proposed method has been established as a more generic one in its applicability for mining anomalous sub-graphs in graph representation of network data.

Further work in this direction could be on improving the candidate sub-graphs generation procedure and employing a more subjective anomaly identification methodology. One can even try with a different set of parameter values.



(a) Facebook-107 graph (b) AS20000102 graph

Fig. 3. Results of detecting anomalous sub-graphs on benchmark graph data sets

Acknowledgment. The authors would like to thank Director, CAIR for supporting this work. The authors also thank Dr. L. Akoglu for providing the source code of her paper [1].

References

1. Akoglu, L., McGlohon, M., Faloutsos, C.: Oddball: Spotting anomalies in weighted graphs. In: Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V. (eds.) PAKDD 2010. LNCS, vol. 6119, pp. 410–421. Springer, Heidelberg (2010)
2. Albanese, A., Pal, S.K., Petrosino, A.: Rough sets, kernel set and spatio-temporal outlier detection. IEEE Trans. on Knowledge and Data Engineering (2012) (online)
3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Computing Surveys 41(3), 15.1–15.58 (2009)
4. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. PNAS 99(12), 7821–7826 (2002)
5. Lee, D.D., Seung, H.S.: Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791 (1999)
6. Leskovec, J.: Stanford network analysis platform, SNAP (2013), `http://snap.stanford.edu/data/index.html`
7. McAuley, J., Leskovec, J.: Learning to discover social circles in ego networks. In: NIPS, Nevada, USA, pp. 548–556 (2012)
8. Noble, C.C., Cook, D.J.: Graph-based anomaly detection. In: Proc. SIGKDD, Washington, DC, USA, pp. 631–636 (August 2003)
9. Rattigan, M.J., Jensen, D.: The case for anomalous link discovery. SIGKDD Explorations 7(2), 41–47 (2006)
10. Suri, N.N.R.R., Murty, M.N., Athithan, G.: Data mining techniques for outlier detection. In: Zhang, Q., Segall, R.S., Cao, M. (eds.) Visual Analytics and Interactive Technologies: Data, Text and Web Mining Applications, ch. 2, pp. 22–38. IGI Global, New York (2011)
11. Wang, F., Li, T., Wang, X., Zhu, S., Ding, C.: Community discovery using nonnegative matrix factorization. DMKD 22(3), 493–521 (2011)

Discriminant Analysis for Identifying Individuals of Electrocardiogram

Yogendra Narain Singh

Institute of Engineering & Technology,
Gautam Buddh Technical University Lucknow
singhyn@gmail.com

Abstract. We present a novel method to characterize the electrocardiogram (ECG) for individual recognition. The method works on analytical and appearance features of the heartbeats. The features are analyzed using the principle of Fisher's linear discriminant that produces well separated classes in a lower dimension subspace, under the presence of noise and muscle flexure. The biometric experiment is benchmarked using ECG recordings of the publically available QT database. The proposed method achieves the recognition accuracy of 98.9% on the evaluated subjects which is found optimum than the other best known methods.

Keywords: Identification, Electrocardiogram, Heartbeats, Biometrics.

1 Introduction

Many body parts, personal characteristics and signaling methods have been suggested to use as biometrics for identifying individuals. The biometric attributes that are secured and robust against falsification are the major concerns of the biometric research. The reason that the biometric attributes are not robust against falsification because they hardly ensure the liveness property [1]. The search of a new biometric modality that has an inherent property of liveliness is of greater interest. In recent years, the physiological signal especially the electrocardiogram (ECG) has been proposed a novel biometric for human recognition [2]-[5].

An ECG is the graphical recording of the cardiac electric potential as a function of time. The ECG signal is persistent and repetitive. The ECG is generated from a complex self-regulatory physiological processes which makes it highly confidential and difficult to spoof. The ECG signal is an ultimate liveness metric and acceptable due to the latest advances in the sensing technologies [6]. In multimodal framework the ECG can be combined with face and fingerprint biometrics for secure and robust individual recognition [7]. However, the ECG signal as a biometrics has notable advantages but it has limitations that must be addressed before the operational deployment of ECG biometric system [8]. The ECG data contains a relatively low information content due to low signal-to-noise ratio. The ECG signal varies with an individual emotion, excitement and stress level. There is a need of standard data representation technique and a careful transformation of signal features that preserve the discriminative information between subjects.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 94–99, 2013.

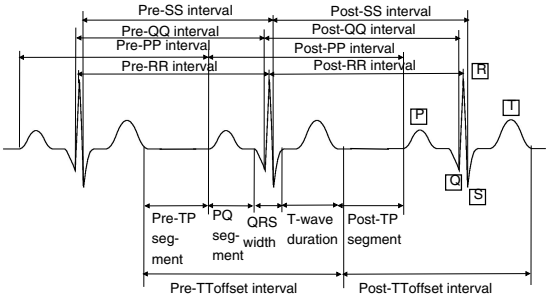


Fig. 1. Extraction of interval features from heartbeats

This paper analyzes the ECG using a novel method that is insensitive to signal variations. The method utilizes a hybrid approach consisting analytical and appearance methods of ECG characterization. The fiducial points of interest are extracted from ECG waveforms using analytical method whereas the appearance method extracts the morphological features of the heartbeats. In the ECG analysis, the analytical features are used as a complementary information with the appearance features. The separability between inter-subject ECG signals and the similarity between intra-subject ECG signals can be drawn using discriminant analysis [9] and the subject classification is performed using a nearest neighbor classifier. The method is benchmarked using ECG recordings selected from publically available QT database [10]. In the sections to follow the methodology and the experimental setup of ECG biometric system along to the results of people recognition are presented.

2 Linear Discriminant Analysis of ECG Signals

This section proposes a computational scheme for evaluating the metric of different ECG signal attributes for their discrimination potential. A normal heartbeat consists of three main components: P wave, QRS complex, and T wave (see for instance Fig. 1). The heartbeats are detected using QRS complex delineator [11]. It uses digital analysis of slope, amplitude and width information of ECG waveforms. After segmenting the heartbeats, the P wave delineator [12] and the T wave delineator [13] are employed to determine the P wave and the T wave along to their end fiducials, respectively. We propose a novel framework of ECG characterization that extracts salient features from each heartbeat, i.e., interval features and morphological features.

The heartbeat interval features contains, the QRS width, the T wave duration, the PQ segment, the pre-TP segment and the post-TP segment as shown in Fig. 1. In this class of features the temporal duration between PP, QQ, SS, TT and RR waves of successive heartbeats are also extracted. The pre-PP (post-PP) interval is the time interval between P_{onset} of a given heartbeat and the P_{onset} of the previous (following) heartbeat. The pre-QQ (post-QQ) interval is the

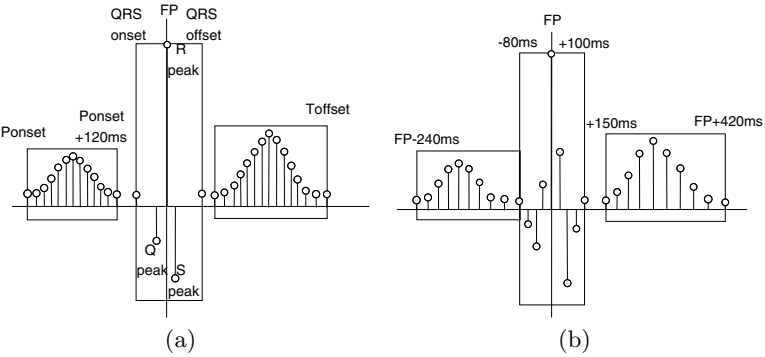


Fig. 2. Extraction of morphological features from heartbeat (a) normal and (b) scaled signals

time interval between Q_{Peak} of a given heartbeat and the Q_{peak} of the previous (following) heartbeat. The pre-SS (post-SS) interval is the time interval between S_{peak} of a given heartbeat and the S_{peak} of the previous (following) heartbeat. The pre-TT (post-TT) offset interval is the time interval between T_{offset} of a given heartbeat and the T_{offset} of the previous (following) heartbeat. Similarly, the pre-RR (post-RR) interval is defined as the RR interval between a given heartbeat and the previous (following) heartbeat. The heartbeat interval features and the interbeat interval features are shown in Fig. 1.

The morphological features contains the amplitude values of the segmented heartbeats. They are extracted from normal and scaled ECG signal whereas the signal is scaled using Pareto method [14] with an aim to reduce the sensitivity of the ECG signal both to noise and motion artifact. Thirty-two features and twenty-eight features are extracted from temporal windows that are set heuristically from the normal and scaled ECG signal, respectively using linear interpolation of uniformly distributed sample positions as shown in Fig. 2.

A gallery set composed of a relatively large group of subjects of different ECG characteristics is prepared. The database contains several examples of ECG signals for each subject in the gallery set and at least one example in the probe set. We employ linear discriminant analysis (LDA) method that uses class membership information and allows us to find eigenfeatures and therefore representations in which the variations among different ECG signals are emphasized, while the variations of the same ECG signal that are due to different physiological conditions are de-emphasized.

Let S be the representation of an original signal to be analyzed, we construct the lexicographic vector expansion R^n where n is the number of features extracted from the signal. Let us define all instances of an individual ECG as being in one class and the ECGs of different subjects as being in different classes for all subjects (N) in the gallery set. We establish a framework for performing a cluster separation analysis in the feature space. Also, having labeled all

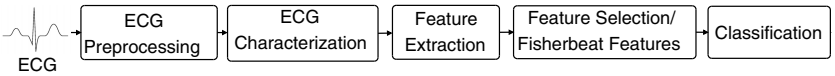


Fig. 3. Classical structure of an ECG biometric recognition system

instances in the training set and having defined all the classes, we compute the within-class and between-class scatter matrices as follows:

$$S_W{}^{(S)} = \sum_{i=1}^{N} \sum_{x_{ij} \in N_i} (x_{ij} - \bar{x_i})(x_{ij} - \bar{x_i})^T \text{ and } S_B{}^{(S)} = \sum_{i=1}^{N} N_i(\bar{x_i} - \bar{x})(\bar{x_i} - \bar{x})^T$$

where $\bar{x_i} = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$ is the mean vector of the class X_i and $\bar{x} \in R^n$ is the mean vector of all samples. The discriminatory power of a representation can be quantified using the separation matrix. The class separation matrix and a measure of separability, $D^{(S)}$ can be estimated by maximizing the expression, $\arg\max_{\phi} \frac{|\phi^T S_B \phi|}{|\phi^T S_W \phi|}$ and derive a set of m feature basis vectors $\{\phi\}_{k=1}^{m}$. We assume that the within-class scatter matrix S_W is non-singular, then the FLD finds ϕ as the k most significant generalized eigenvectors of $S_W{}^{-1} S_B$ those correspond to the first k largest generalized eigenvalues, thus $D^{(S)} = S_W^{-1} S_B \phi_k = \lambda_k \phi_k$. The generalized eigenvectors form the basis representation of the gallery and the probe ECG signals. For identity recognition the classification is performed using a nearest neighbour classifier.

3 Recognition Results of ECG Biometric Method

The classical structure of an ECG biometric recognition systems is shown in Fig. 3. The ECG recordings are first preprocessed that makes necessary correction of signal from noise and muscle flexure. The processed signal is then passed for ECG characterization followed by feature extraction. The Fisherbeat features are selected from heartbeat features and thus most discriminatory information is preserved. Finally the subject classification is performed in the projected domain using nearest neighbor criterion.

The performance of the ECG biometric system is tested on the ECG recordings of QT database. The first channel ECG recordings of 85 subjects are used in this experiment. The gallery and the probe datasets are prepared from ECG recorded at different time segments. The results from the subjects class specific linear projection of the ECG features demonstrate that there exist a correlation between number of Fisherbeat features and subject classification performance. The decrease in intra-subject variability and an increase of inter-subject separability, represented by first few principal components are shown in Fig. 4(a) and Fig. 4(b).

The receiver operating characteristic (ROC) curve related to better recognition rate achievable by the system, drawn for different projected dimensions are shown in Fig. 4(c). The system achieves better result at dimension ten (DIM 10) such as the values of the GAR are reported to 96% at zero FAR and 100% at 2% of FAR. The performance of the system degrades at higher dimensions and also for the dimensions below ten.

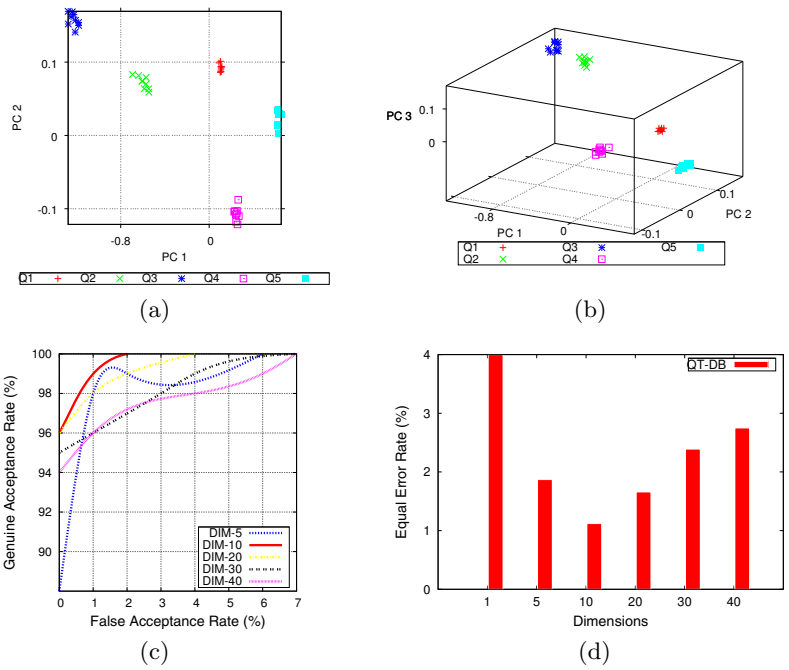


Fig. 4. Discriminability between ECG signals of five subjects represented by first (a) two and (b) three principal components. (c) ROC curve for different projected dimensions. (d) The histogram representation of recognition performance of discriminant features.

The performance of the Fisherbeat features vary with the number of principal components. The equal error rate (EER) values gradually decrease with an increase in component dimensions and found lowest to 1.1% at dimension ten. For higher dimensions of principal components, EER values raise very steadily. This trend in performance is shown in Fig. 4(d) by histogram drawn between EER and dimensions. These results signify that the discriminatory information of an ECG signal has been significantly captured by the Fisherbeat features. In particular, the Fisher-eigenspace of significantly lower dimension is capable to represent the discriminatory information of an individual heartbeat.

4 Conclusion

This study has proposed a novel method to characterize the ECG signal for individual recognition. The features of classes heartbeat interval features and waveform morphological features are derived. Using discriminant analysis the Fisherbeat features are selected. The advantages of using the discriminant analysis of the heartbeat features are the suppression of noise from the signal, reducing the complexity to access a larger attribute set and simplifying the classification process. The reported results have proved the effectiveness of the proposed ECG biometric method for individual recognition.

References

1. Singh, Y.N., Singh, S.K.: A Taxonomy of Biometric System Vulnerabilities and Defenses. Intl. J. of Biometrics 5(2), 137–159 (2013)
2. Biel, L., Pettersson, O., Philipson, L.: ECG Analysis: A New Approach in Human Identification. IEEE Trans. on Instrumentation and Measurement 50(3), 808–812 (2001)
3. Singh, Y.N., Gupta, P.: ECG to Individual Identification. In: Proc. Biometrics: Theory, Applications and Systems, Washington DC, USA, pp. 1–8 (2008)
4. Singh, Y.N., Gupta, P.: Biometrics Method for Human Identification Using Electrocardiogram. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1270–1279. Springer, Heidelberg (2009)
5. Singh, Y.N., Gupta, P.: Correlation Based Classification of Heartbeats for Individual Identification. J. Soft Computing 15(3), 449–460 (2011)
6. Zhao, C., Wysocki, T., Agrafioti, F., Hatzinakos, D.: Securing Handheld Devices and Fingerprint Readers with ECG Biometrics. In: Proc. IEEE Fifth Int. Conf. Biometrics: Theory, Applications and Systems, pp. 150–155 (2012)
7. Singh, Y.N., Singh, S.K., Gupta, P.: Fusion of Electrocardiogram with Unobtrusive Biometrics: An Efficient Individual Authentication System. Pattern Recognition Letters 33, 1932–1941 (2012)
8. Singh, Y.N., Singh, S.K.: Evaluation of Electrocardiogram for Biometric Authentication. Journal of Information Security 3, 39–48 (2012)
9. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Wiley India, New Delhi (2009)
10. Laguna, P., Mark, R.G., Goldberger, A., Moody, G.B.: A Database for Evaluation of Algorithms for Measurement of QT and Other Waveform Intervals in the ECG. Comput. Cardiol. 24, 673–676 (1997)
11. Pan, J., Tompkins, W.J.: A Real Time QRS Detection Algorithm. IEEE Trans. on Biomedical Engineering 33(3), 230–236 (1985)
12. Singh, Y.N., Gupta, P.: A Robust Delineation Approach of Electrocardiographic P Waves. In: Proc. 2009 IEEE Symposium on Industrial Electronics and Applications, Kuala Lumpur, Malaysia, vol. 2, pp. 846–849 (2009)
13. Singh, Y.N., Gupta, P.: A Robust and Efficient Technique of T Wave Delineation from Electrocardiogram. In: Proc. Second Int. Conf. on Bio-inspired Systems and Signal Processing, Porto, Portugal, pp. 146–154 (2009)
14. van den Berg, R.A., Hoefsloot, H.C.J., Westerhuis, J.A., Smilde, A.K., van der Werf, M.J.: Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data. BMC Genomics 7(142), 1–15 (2006)

Trust Based Secure and Energy Efficient Clustering in Wireless Sensor Network: A Bee Mating Approach

Rashmi Ranjan Sahoo[1], Abdur Rahaman Sardar[2], Moutushi Singh[3], Sudhabindu Ray[1], and Subir Kumar Sarkar[1]

[1] Department of ETCE, Jadavpur University, Kolkata, India
[2] Department of CSE, NITMAS, West Bengal, India
[3] Department of IT, IEM, Kolkata, India
{rashmi.cs2005,abdur.sardar,moutushisingh01}@gmail.com,
sudhabin@etce.jdvu.ac.in, su_sircir@yahoo.co.in

Abstract. In this paper we have proposed a trust based secure and energy efficient clustering algorithm in wireless sensor network using Honey Bee Mating Algorithm (TBCR-BMA). The proposed TBCR-BMA deprives the malicious node to act as cluster head, thus prolong the life time of the network. Moreover, we reveal that this proposed scheme outperforms the most popular hierarchical Low Energy Adaptive Clustering Hierarchy (LEACH) and Advertisement timeout driven bee's mating approach to maintain fair energy level in sensor networks (TBCMA) in terms of average residual energy of nodes and total energy of the network.

Keywords: Wireless Sensor Network, Clustering, Trust, HBMA.

1 Introduction

Energy consumption is one the foremost factor for wireless sensor network (WSN), which is inversely proportional to the life span of the WSN. By clustering the nodes and reducing the number of transmission, its lifetime is enhanced [1]. In our proposed work we have considered a trust based secure and energy efficient clustering algorithm using Honey Bee Mating Algorithm (HBMA) inspired by the conduct of social insects. HBMA is used to find the most appropriate cluster head (CH) in WSN. The clustering architecture of WSN localizes the route set up within the clusters, reduces size of the routing table of individual sensor nodes, eliminate the transmission redundancy and data aggregation from multiple nodes [2–4]. We have calculated trust value on each node for ensuring selected CH_s trustworthiness and highest remaining energy.

2 Proposed Honey Bee Mating Based Clustering

This section presents the proposed clustering method. The operation of the protocol is broken up into cluster head(CH) selection and cluster creation process.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 100–107, 2013.

Further CH section consists of premiere phase and steady-state phase. Details procedure for finding the trustworthiness of a node can be found in section 3.

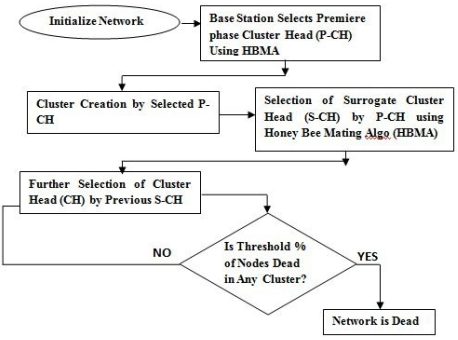


Fig. 1. Flowchart of artificial honey bee mating based clustering

2.1 Cluster Head Selection Process

The mechanisim of CH election mechanism is carried in two different phases. In first phase (Premiere Phase) the entire CH election process is govern by base station (BS) where as in subsequent phases (Steady Phase) the election is carried by the CH of previous phase.

2.1.1 Premiere Phase Secure Cluster Head Election Process

In this phase, as the entire CH selection process will be carried by BS so there will be less over head on different sensors node in terms of energy consumptions.

Algorithm Premiere Phase Secure Cluster Head Election Process

1. BS broadcast a $START - ELECTION$ message to find out the expected premiere Phase cluster head (CH_{EPP}).
2. The node which wishes to be a candidate for CH will send a *WISH_FOR_CH* packet to BS with its Node id (N_{id}), Residual Energy (R_E), Distance from BS (D) and Neighbor node list present within its transmission range.
3. BS multicast a message to all the neighbor of each CH_{EPP} for Direct Trust (DT_{EPP-CH}) [Discussed in section 3].
4. BS compares the direct trust of each individual expected cluster head with supplied trust thresh hold value, if BS found $DT_{EPP-CH} > T_{th}$ then consider that node as expected premiere Phase cluster head $(CH_{EPP})list$.
5. BS executes Honey Bee Mating Algorithm (HBMA) on expected premiere Phase cluster head (CH_{EPP}) list by preparing chromosome structure1 (Discussed in section 2.3) and finds the $P - CH : CALL\_HBMA()$
6. BS station declare that node as premiere phase cluster head ($P - CH$) and broadcast a message PCH_ADV_MSG with N_{id} of CH across the network

2.2 Cluster Formation Process

In cluster formation, cluster size is very important for network lifespan. With a big cluster size and distant between CH and BS, the overhead for data transmission is more; hence energy consumption by that CH will be more as compared to large cluster with small distance from BS. Same thing happens for small cluster with large distance from BS than small cluster with small distance from BS. So our proposed cluster formation process considers the equipoise size of cluster. For balancing the CH load across the network, cluster size (SC) is considered the as a function of distance between CH and BS i.e.$SC = f(D(CH, BS))$. When the $D(CH, BS)$ is large the corresponding CH will accommodate small number of nodes. Hence our proposed method ensures stable cluster in terms of load and energy dissipation across the network. The entire cluster creation process is as follows:

1. Every recipient of $P_{CH\_ADV\_MSG}$ message expect P-CH decides which cluster it will Join as a cluster member depending upon its own R_E and distance from premiere cluster head.
2. Every non-cluster member calculates value of $P-CH$ as ($VP-CH$), present within its radio range in order to join as cluster member. $VP-CH_i = W_1 * RE_i + W_2 * D(P-CH_i, BS) + W_3 * D(P-CH_i, Nc_i)$, where $D(P-CH_i, BS)$, is the distance between $P-CH$ and BS. $D(P-CH, NC_i)$ is the distance between $P-CH$ and non-cluster member.W_1, W_2, W_3 are different weights such that $W_1 + W_2 + W_3 = 1$.
3. Each non-cluster members decides $P-CH$ depending upon the minimum value of $VP-CH_i$ and uses CSMA to inform $P-CH$ which it belongs.
4. After $P-CHs$ received messages from all nodes, it will create Time Division Multiple Access(TDMA) scheduling table and send it to all nodes, which contains the time allocated to each node for transmitting data to the $P-CH$.
5. To avoid the collision $P-CHs$ will issue new TDMA slots to all nodes in their clusters when thresh hold allocated time elapsed.

2.2.1 Steady Phase Cluster Head Selection

After the selection of $P-CH$ by BS and cluster formation by $P-CH$, it elects a surrogate cluster head ($CH_{surrogate}$) as spare CH before it run out a thresh hold energy. Also, $CH_{surrogate}$ finds another CH before its energy level reaches to a certain threshold level and so on. This procedure will continue until a certain percentage of sensor nodes run out their energy completely within any cluster. The detail description of $CH_{surrogate}$ election is as given below.

1. Cluster Head selected at premiere phase (act as CH) compares its own residual energy level with supplied Residual Energy Threshold (TH_{R-E}) value. When the residual energy level reaches at supplied TH_{R-E} value at that point of time, $P-CH$ starts the election of $CH_{surrogate}$.
2. CH broadcast a message within its cluster members to get the R_E, TRUST (Direct and Indirect) and cluster head counter (CH_C) value.

3. CH compares the TRUST value of each cluster members with supplied total trust threshold (Th_{total_trust}) value, if TRUST value of each individual member is greater than the Th_{total_trust} then that node will be considered as a candidate for $CH_{surrogate}$ otherwise will be considered as malicious one.
4. CH execute the Honey Bee Mating Algorithm (HBMA) on surrogate cluster head list by preparing chromosome structure 2 (Discussed in section 2.3), in order to select the $CH_{surrogate} : CALL\_HBMA()$

2.3 Problem Mapping

In this section we have modeled the proposed clustering algorithm of wireless sensor network as per the honey bee mating algorithm. Various controlling parameters of HBMA are mapped with the proposed clustering algorithm are as follows The parameters of real honey bee mating such as Nodes in the Network, Random selection of Expected cluster Head, Queen, Drone Bee, Worker Bee and Mating are mapped onto Bees in hive, Initial Population (Chromosome), Best Bee (Selected by Fitness Function), Expected cluster head list (initial population) - Best bee (Queen), Heuristic Search Function and Cross Over respectively of artificial bee.

- Initial population is a set of random possible solution comprise of chromosome. In this work, we have considered two different type of chromosome structure, one structure for premier phase cluster head selection and other for steady phase.
- Premier Phase Chromosome
  The chromosome structure C_i for premier phase are consists of series of n genes.
  $C_i = \{g_{i1}, g_{i2}, g_{i3} \cdots g_{ij} \cdots g_{ik}\}$ where i=1, 2,3... population size and
  $g_{ij} = \begin{cases} 1, & \forall Premierphaseexpectedclusterhead \\ 0, & \forall Regularnode \end{cases}$
  g_{ij} stands for node n_i associated with chromosome c_i (Shown in Fig. 2.)
- Steady Phase Chromosome
  Fig. 3. shows the structure of steady phase chromosome. First row of the chromosome structure 2 represents node IDs and the values of respective cells are their cluster head ID. In this structure, those nodes which are chosen as cluster head have 0 in their particular cells. Nodes which are not a member of the corresponding cluster head are filled with -1. For example node with id 3, 6 and 8 are not member of cluster head 5.
- Fitness Function
  In this work we have designed two different fitness functions (one for premier and other for steady phase cluster head selection) by considering all the parameters of the problem that affects the fitness of an individual chromosome. The fitness function for premier phase cluster head selection depends on the Residual Energy (R_E), Direct Trust by neighbor node of expected cluster heads (DT_{EPP-CH}) , Distance between expected cluster heads and Base station ($D_{EPP-CH,BS}$).

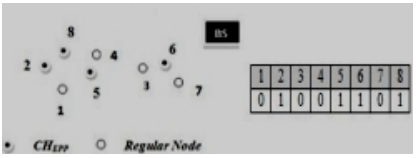


Fig. 2. Chromosome structure 1

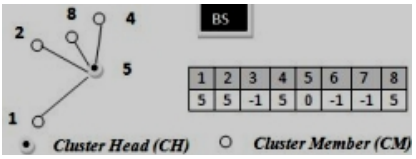


Fig. 3. Chromosome structure 2

$$fitnes(chrome1_k) = \alpha * \sum_{i=1}^{n} R_{EEPP-CH_i} + \beta * \sum_{i=1}^{n} DT_{EPP-CH_i} + \gamma * \frac{1}{\sum_{i=1}^{n} Dist(BS, EPP - CH_i)} \quad (1)$$

Where $fitnes(chrome1_k)$ denotes k_{th} chromosome of chromosome structure 1 and α, β, γ are weighted coefficients, $\alpha + \beta + \gamma = 1$.

During steady phase surrogate cluster head selection following fitness function is used for evaluating the best individual chromosome from chromosome structure 2.

$$fitnes(chrome2_k) = \alpha * \sum_{i=1}^{n} R_{ECM_i} + \beta * \sum_{i=1}^{n} IT_{(CM_i,CH)} + \gamma * \sum_{i=1}^{n} DT_{(CM_i,CH)} + \eta * \frac{1}{\sum_{i=1}^{n} Dist(CH, CM_i)} \quad (2)$$

Where $fitnes(chrome2_k)$ denotes k_{th} chromosome of chromosome structure 2. R_{ECM} is the residual energy of cluster member. $IT_{(CM,CH)}$, $DT_{(CM,CH)}$ are the indirect and direct trust of cluster members calculated by the CH respectively. $Dist(CH, CM)$ is the distance between CH and cluster members.

```
Procedure CALL_HBMA ( )
Initialization
Initialize all the bees (initial population);
Evaluate the fitness (fitness_i) of the population;
Select the bee with highest fitness as Queen and
set fitness_Queen = fitness_Best_Bee;
for Iter=1: Maxiter do
begin
    while  Spermathica_Queen is NOT full do
    begin
        Select a drone depending on depending on
        Prob (Q, D) as in equation (7);
        Store the drone in queens spermatheca;
    end while
for i=1 to no. of brood do
```

```
begin
    Broodi = Dronei + rand (0, 1)*(Queen-Dronei);
    Improve the broods by local search;
end for
Select the brood with highest fitness as best_brood;
if fitness_best_brood  > fitness_Queen then
Q = fitness_best_brood ; %Replace the Queen with best brood
end if
end for
```

3 Trust Evaluations of Sensor Nodes

Trust value of a sensor node is calculated from the history of transactions of the node and from the recommendations given by other neighbor node of the cluster to ensure the sensor nodes trustworthyness. Trust value is the level of confidence of a node N_i on neighbor node N_j depending on the performance of the assigned task [7, 8]. To find trust each sensor node keeps track of the conduct of their neighbors and maintains record of various parameters called as trust metrics. Initially, when a sensor node joins the network, it is assumed that the node N is a Trustworthy (Benevolent) and some trust value is allocated to the node depending upon the threshold trust value (T_{th}). In our proposed scheme Tth is 0.5 [7, 8]. Trust is time dependent. For each of the successful transaction between node N_i and N_j the trust metrics value of corresponding node will increase up to maximum 1, otherwise it will decrease up to as minimum as 0. We considered various trust metrics like Data Packets Forwarded (DPF), Control Packet Forwarded (CPF), Data Packet/message Precession (DPP), Control Packet/message Precession (CPP), Packet Address Modification (PAM), Injection of false Packet ($IOFP$) to measure the trust worthiness of nodes. Direct Trust (DT) of any node is geometric mean of all the trust metrics i.e. DT of node N_i on node N_j can be calculated as shown in equation (3).

$$DT_{Ni}(N_j) = [\prod_{k}(m_k)]^{\frac{1}{k}}. \tag{3}$$

Where $DT_{Ni}(N_j)$ represents the direct trust of node N_i on node N_j and m_k is set of k different trust metrics.

$$IT_{Ni}(N_j) = [\prod_{n}(DT_n(N_j))]^{\frac{1}{n}}. \tag{4}$$

Where $IT_{Ni}(Nj)$ is the indirect trust of node N_i on sensor node N_j, calculated from indirectly given information (i.e. Direct Trust (DT) on N_j) by n neighbor and given to cluster-head (CH).

$$TT_{Ni}(N_j) = W_1 * \sum_{i=1}^{n}(\sum_{j=1}^{n} DT_{Ni}(N_j)) + W_2 * \sum_{i=1}^{n}(\sum_{j=1}^{n} IT_{Ni}(N_j)) \tag{5}$$

Where $TT_{N_i}(N_j)$ is the total trust of node N_i on N_j. Where W_1 and W_2 are the weightage given to DT and IT depending on the applications.

4 Simulation Result and Discussion

Simulation work has been carried in MATLAB and some malicious nodes has been injected intentionally to ascertain the robustness of proposed TBCR-BMA.

Test Bed Setting: Test bed dimension: $500m * 500m$, Initial node deployment: Random, No. of nodes: 150, Node movement: Static, Radio range of node: 30 m, Radio range of base station: Full Network, Sink location: $500m * 500m$, No. of trust metrics: 6, Initial trust metric value: 0.5, Trust value range: 0.0 to 1.0, Broadcast packet size: 25 Byte, Data packet size: 50 Byte, Initial node energy: 3Joule, Transmitter/Receiver circuitry dissipation: 0.5 nJ / bit, Data aggregation Energy (EDA):0.5nJ /bit, Cluster Head Energy Consumption: 0. 7 nJ / bit.

HBMA Parameter Setting: Number of population: 30, No. of drones: 29, No. of broods: 29, Capacity of Spermatheca: 29.

Figure 4 shows the premier phase cluster head selection along with the cluster creation by base station.

A. Alive Node $fig.5$ depicts the percentage of live nodes in the network over 250 simulation rounds.In proposed TBCR-BMA the number of node drained out their energy is less (i.e. more no. of live nodes) as compared to LEACH and TCBMA [5, 6].

B. Average Residual Energy Average remaining energy of nodes in cluster in terms of joules is shown in $fig.6$ over 250 simulation rounds. Here the cluster formation is done only once, so each CH and its members consume less energy as compared to other two methods [5, 6]. Hence our proposed method leaves the CH and its member with more energy.

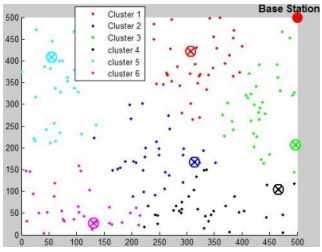


Fig. 4. $P - CH$ selection by base station and cluster creation

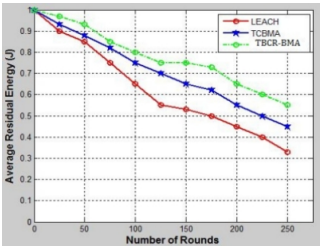


Fig. 5. No. of Alive Nodes

C. Total Energy Consumption From $fig.7$ it is clear that proposed TBCR-BMA has consumed less energy as it deprives the malicious node to become a CH. Also, TBCR-BMA forms the cluster only once. So, TBCR-BMA performs better in 250 rounds as compared to LEACH [5] and TCBMA [6].

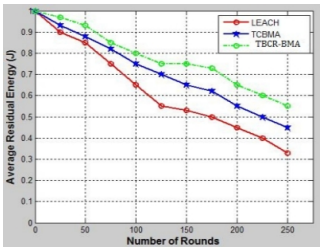


Fig. 6. Average Residual Energy

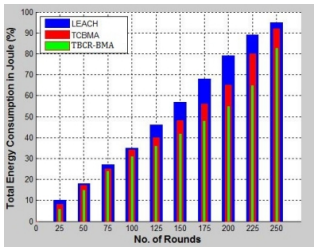


Fig. 7. Total Energy Consumption

5 Conclusion

We have implemented TRUST mechanism to select a benign CH. Malicious node may inject false packet, change destination address or drop data packets. This benign CH will not absorb any harmful activities and prolongs network life time. Furthermore, for balancing the energy consumption among CH_s, we also ensure that the clusters closer to the BS have smaller sizes than those farther away from it. Thus CH_s closer to the BS can also preserve some energy.

References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless Sensor Network Survey. Computer Networks 52(12), 2292–2330 (2008)
2. Liu, X.: A Survey on Clustering Routing Protocols in Wireless Sensor Networks. Sensors, 11113–11153 (2012)
3. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. Computer Commununication 30, 2826–2841 (2007)
4. Rajagopalan, R., Varshney, P.K.: Data-aggregation techniques in sensor networks: A survey. IEEE Communication Survey Tutorial 8, 48–63 (2006)
5. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy- Efficient Communication Protocol for Wireless Microsensor Networks. In: IEEE Proceedings of the Hawaii International Conference on System Sciences, pp. 1–10 (January 2000)
6. Senthilkumar, J., Chandrasekaran, M., Suresh, Y., Arumugam, S., Mohanraj, V.: Advertisement timeout driven bee's mating approach to maintain fair energy level in sensor networks. Applied Soft Computing 12(7), 1884–1890 (2012)
7. Sahoo, R.R., Panda, R., Behera, D.K., Naskar, M.K.: A trust based clustering with Ant Colony Routing in VANET. In: Proceedings of IEEE International Conference on Computing Communication & Networking (ICCCNT), pp. 1–8 (July 2012)
8. Sahoo, R.R., Singh, M., Sardar, A.R., Mohapatra, S., Sarkar, S.K.: TREE-CR: Trust based secure and energy efficient clustering in WSN. In: Proceedings of IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), March 25-26, pp. 532–538 (2013)

Training by ART-2 and Classification of Ballistic Missiles Using Hidden Markov Model

Upendra Kumar Singh[1] and Vineet Padmanabhan[2]

[1] PGAD, Defence Research & Development Organisation, Hyderabad, India
singh_ukin@yahoo.co.in
[2] School of Computer & Information Sciences, University of Hyderabad, India
vineetcs@uohyd.ernet.in

Abstract. This paper addresses the classification of different ranges of Ballistic Missiles (BM) for air defense applications using Adaptive Resonance Theory (ART-2) and Hidden Markov Model (HMM). ART-2 finds the initial clusters using unsupervised learning to be fed to HMM for classification using recursive method. The classification is based on derived parameters of specific energy, acceleration, altitude and velocity which in turn are acquired from measured data by radars. To meet the conflicting requirements of classifying short as well as long-range BM trajectories, we are proposing a formulation for partitioning the trajectory by using a moving window concept. Experimental results show that the HMM model is able to classify above 95% within time of the order of milliseconds once initial data is trained using ART2.

1 Introduction

In this work, we deal with the classification of ballistic missiles in real-time for air-defense application. The classification of ballistic missile trajectory is a challenging problem due to its time-varying dynamics and short response time available for interception. The general policy of Ballistic missile defense is to neutralize incoming threats at higher height and at longer range so that debris falls away from intended impact zone [8]. This may not be realized due to late detection and short response time of launch. In a typical ballistic missile defense (BMD) scenario the radar first detects the target missile and information is communicated to the launcher. The launcher in-turn classifies the target to decide about the time of launch and height of interception. The response time is dependent on many factors including range of radars for acquiring target of the given radar cross section (RCS), target velocity, preparation time of interceptor, velocity of interceptor, time of flight of interceptor, planned height, range of kill and classification time. Classification of variable-length trajectory and time varying dynamic attributes, which are important ingredients for ballistic missile classification, cannot be done using conventional techniques like neural networks, Bayesian or Kernel methods [2]. Classification of motion trajectories using Hidden Markov Models (HMMs) is presented in [9] and Gaussian mixture-based HMMs for trajectory modeling and classification are described by Bashir,

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 108–114, 2013.

et.al [1]. In this paper a novel procedure is formulated to group initial data using unsupervised ART-2 algorithm and then classify ballistic missiles using HMM.

2 Problem Formulation

The radar measures only the position of the target and this information is noisy [11]. The kalman filter estimates position, velocity and acceleration by processing the noisy measured position data. The filtered kinematic parameters of missiles are passed to linear quantizer to convert the data into discrete forms before it passes through HMM. For the present work, we are using 4 kinematic parameters, which are computed based on the kalman filter estimates. These 4 parameters are (a) Specific energy, (b) Acceleration, (c) Altitude, and (d) Velocity. Also the number of classes considered are three, namely M400 Class of Ballistic Missile (400 Km range), M1000 Class of Ballistic Missile (1000 Km range) and M2000 Class of Ballistic Missile (2000 Km range). HMM parameters are initialised, and Baum-Welch (BW) algorithm is applied to get a new set of parameters of HMM that has maximum likelihood of occurrence, given a set of observed feature vectors. Running many iterations of BW yields a sequence that approaches a local maximum of the likelihood. The Viterbi algorithm finds the most probable way through the model Λ and evaluate the model by maximising probability of correspondence with a trajectory class. Hidden Markov Model (HMM) is trained for a set of points of the trajectory and evaluated to assess the performance of the model using trajectories of ballistic missiles. The training time taken for computation is less than 100 ms and it is highly suited for applications where scene of interest is changing continuously and next state vector is dependent on only a limited last states within a set window. However, initial set of data needed for HMM has to be pre-specified or known using another algorithm. ART-2, a competitive learning algorithm, is used here for self-organising input pattern based on user defined parameters of the network.

Since ballistic missiles can be programmed to fly in nominal, lofted and depressed mode the parameters used for classification varies for same as well as different classes of target missiles. To meet the conflicting requirements of classifying short as well as long range trajectories, a formulation is proposed for partitioning the trajectory by using moving window concept. This concept allows us to use parameters in localized frame which helps in reducing the problem to fit into the same model. This moving window concept of testing and training brings out the knowledge embedded in the system and deals with variable length inputs. The length of the moving window is selected such that it fully captures the dynamics of the target which are intended to be classified.

3 Adaptive Resonance Theory (ART2)

An adaptive resonance theory (ART) [5] involves three groups of neurons: input processing units (F1 layer), cluster units (F2 layer), and reset units. The F1 layer consists of six types of units (W, X, U, V, P and Q). There are ‘n‘ (dimension

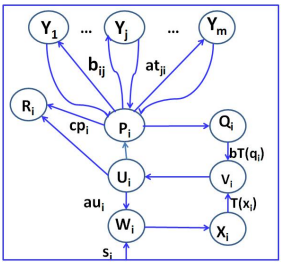


Fig. 1. ART2 Architecture

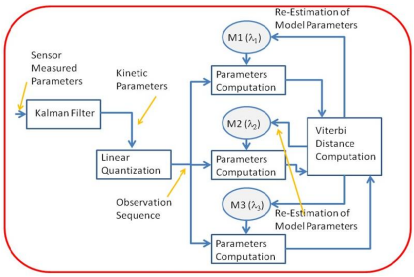


Fig. 2. HMM Recogniser

of input pattern) units of each of these types whereas Figure 1 shows only one unit. S_i forms the input vector. Layer F2 is shown as units $Y_1, \ldots, Y_m$. There are two sets of connections between the layers F1 and F2. The bottom-up weights connecting F1 to F2 are denoted by b_{ij} and the top-down weights connecting F2 to F1 are designated t_{ji}. The cluster unit with the largest net input becomes the candidate to learn the input vector. The criterion for an adequate match between an input pattern and a chosen cluster (feature) is determined by a vigilance parameter, which ranges between 0 and 1. The units V_i and P_i in the F1 layer send signals to the corresponding reset unit R_i. The reset mechanism can check for a reset each time it receives a signal from P_i. The steps involved in ART-2 are (1) As each input is presented to the network, it is compared with the prototype vector that matches most closely. (2) If the match between the prototype and the input vector is NOT adequate, a new prototype is selected. In this way, previous learned memories are not eroded by new learning. The ART2 network is used for classifying initial set of samples to be utilised further for HMM beyond 20 samples. A typical ART2 architecture [3] is re-produced in Figure 1. The parameters chosen for the network for training initial set of values have been fixed after experiment and tabulated in Table 1.

4 Hidden Markov Model (HMM)

HMM is a statistical method which is rich in mathematical structure and forms basis for use in a wide range of applications [9,7,4]. A trajectory is a continuous quantity which can be described as the position of the object in time. HMMs are finite state stochastic machines that robustly model temporal variations in time series data which satisfies the Markovian property [6]. The basic theoretical strength of the HMM is that it combines modeling of stationary stochastic processes and the temporal relationship among the processes together in a well-defined probability space [10]. Considering the trajectory of missile of a particular class to be described at any time as being in one of a set of N distinct states $S_1, S_2, \ldots, S_N$. At regularly spaced discrete times, the system undergoes a change of state according to a set of probabilities associated with the state. Denoting the time instants associated with the state changes as t=1, 2, ..., the actual state at time t is denoted as q_t. Main characteristics of HMM are 1) Given

Table 1. Choice of parameters of ART2

Name of Parameters	**Values**	**Name of parameters**	**Values**
No. of Inputs (F1 Layer)	4	Vigilance Parameter (ρ)	0.999985
No. of Inputs (F2 Layer)	4	Fixed weights in the F1 layer (a, b)	10, 10
No. of Epochs	1	Fixed weight used for reset (c)	0.1
No. of Iterations	200	Activation of winning F2 unit (d)	0.9
Learning rate (α)	0.1	Noise suppression parameter (θ)	$\frac{1}{sqrt(n)}$
Initial top-down weights $t_{ji}\{O\}$	0	Initial bottom-up weights $b_{ij}\{O\}$	$\frac{1}{((1-d)sqrt(n))}$

the $(t-1)^{th}$ state variable, the t^{th} state variable is independent of previous variables and 2) the t^{th} observation depends only on the t^{th} state.

The trajectory classification problem is formulated as to identify the class $C_i (i = 1 \ldots L)$ to which the trajectory state sequence belongs. Each trajectory class C_i is represented by a Model, λ_i , where $\lambda_i = \{\pi, A, B\}$ in which π is initial state probabilities, A is state transition probabilities of dimension (N × N) and B is observation symbol probability distribution of dimension (N × M). Number of states for HMM for Missile classification are considered as N=25; Number of Observation Symbols, M for Specific Energy is 200, Acceleration is 40, Height is 60 and Velocity is 80. Initial state probability, π and state transition probabilities are set to be equal to 1/N. Observation symbol probability, B is based on Gaussian (with mean and variance) representation for the distributions of observations within each state. HMM recognizer as given in Figure 2 shows three HMM models (λ_1,λ_2,λ_3) corresponding to three different types of trajectories belonging to M400, M1000 and M2000 class of missiles. Radar measured data is passed through the Kalman filter as indicated by first block. Observation sequence is linearly quantized and passed through Baum-Welch [2,10] computation to arrive at updated parameters of HMM Model. Viterbi [2,10] distance is used to classify how a new sample is related with trained HMM Model. The steps involved in unified training and testing can be summarized as follows: 1) Define all normal trajectory classes in the scene 2) Define Initial models corresponding to each class 3) The trajectory classes in Captured sequence are marked with 20 samples of data and models are adapted according to them using Baum-Welch Algorithm 4) The output is a new classification model 5) New set of data is evaluated against each HMM model using Viterbi Distance 6) Least Viterbi distance is taken as criteria to assign new data to the corresponding Class 7) New data is added at the end of running window of that class & first data is deleted 8) Training is repeated to adapt the Models 9) Step (5) to (8) is repeated for all new set of data. Termination criteria of BW iteration occurs if maximum

changes for consecutive iteration for all three $\{\pi, A, B\}$ parameters are less than $10e-8$. Re-estimation formulas for$\{\pi, A, B\}$ are given in [10].

5 Results and Analysis

Input data is either taken from radar for live target or generated from mathematical six degree-of-freedom (6-DOF). Sub-modules of 6 DOF are propulsion, gravity, aero-dynamics, atmospheric, missile configuration. External disturbance due to wind and misalignment is also considered in the model. Runga-Kutta order 4 is used for solving the 6 equations of 6DOF i.e. 3 for translational and 3 for rotational motions of the body. If data are taken from simulation, noise is added as per the radar specification to make it near realistic scenario, where azimuth and elevation channels are corrupted with 2 milli-radians and range by 20 m. The measurement data is passed through kalman filter to estimate the target state i.e. position, velocities and acceleration. Based on this estimated data the kinematic parameters, namely specific energy, acceleration, altitude and velocity as mentioned in Section-2 are derived and fed with uniform track rate update of 100ms. Figure 3(a), Figure 3(b) and Figure 4(a) shows Viterbi distance of

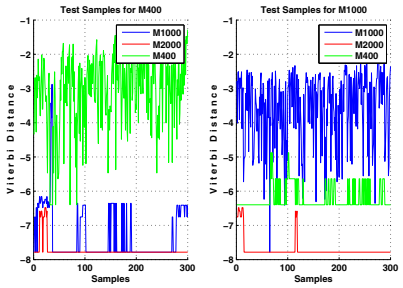


Fig. 3. HMM Results a) M400 b) M1000 for 300 samples for each

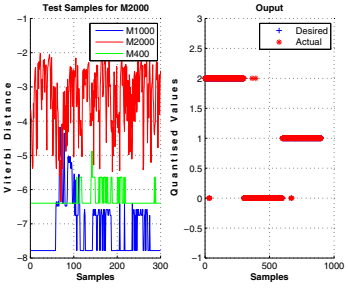


Fig. 4. HMM Results a) M2000 for 300 samples b) Class Category

three classes of missiles as new samples are arriving and being fed to the model for a total of 900 samples for three types of missiles, each with 300 samples. These figures clearly show that trajectory under test are segregated from other two classes for most of the duration. However, at some points, few samples of M1000 trajectory are misclassified as M400 class. Figure 4(b) shows how well desired and computed outputs are matching. Desired symbols for M400, M1000 and M2000 classes are (2,0,1) respectively where a symbol count is incremented during testing if computed values finds matches with correct class. This figure shows that for most of the duration, desired and computed values are overlapping except for very small duration. Results of HMM as recorded in Table-2 demonstrate that pass percentage achieved is more than 95.5% over 900 samples of test data. However, misclassification is ranging from zero to 4.5%. Diagonal values of Table 2 show correct classification of three types of missiles. Off-diagonal values

Table 2. Performance of HMM Model in Confusion Matrix Form

		Actual Class		
		Class 1	Class 2	Class 3
Predicted class	Class 1 (M1000)	298/300	4	5
	Class 2 (M2000)	0	296/300	0
	Class 3 (M400)	2	0	295/300

shows misclassification. HMM takes 1.234 sec for training of 20 samples of initial values in addition to overheads for output storage. HMM takes 44.319 sec of time for unified testing and training of 900 samples which brings 49.353 ms for one sample. The trajectory is validated on a model with different number of states in HMM. HMM with 20 states, 25 numbers of states and above are giving results above 95% whereas performance of the model remains unsatisfactory with 5,10 or 15 states. Time taken by the model for testing new sample of data for 5,10,15,20 and 25 states are 4.816 milli-seconds (ms), 8.02 ms, 18.17 ms, 29.06 ms and 49.24 ms respectively.

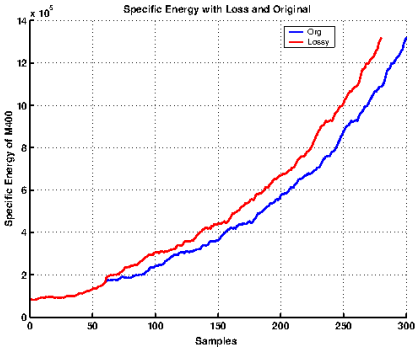


Fig. 5. Trajectory with 2 seconds loss

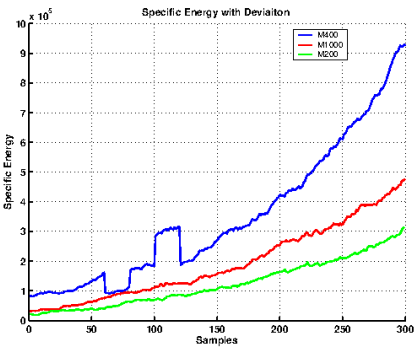


Fig. 6. Trajectory with deviation

The formulation is also tested with two important and practical conditions of losses of data (figure 5) in the track and manoeuvrings targets (figure 6), later being a recent improvement in the ballistic missile development. Figure 5 shows specific energy (SE) plot of M400 class of missile with original and with loss of 20 samples of data equivalent to 2 seconds. The model continues to classify the input data correctly visible as diagonal elements of matrix of Table 3. Numerator indicates model output and denominator denotes total samples passed for validation. Performance of the model is correct classification of 99.31% and incorrect classification for 0.68% for all 3 classes. Results for manoeuvring target are demonstrated as in Table 4. Correct cumulative classification of 97.24% is achieved, whereas incorrect classification amounts to 2.76%. Class 3 for M400 in case of deviated trajectory is passing with 95.29% of correctness.

Table 3. Results with 20 samples Loss

		Actual Class		
		Class 1	Class 2	Class 3
Pred-	M1000	296/300	0	0
icted	M2000	4	298/300	0
class	M400	0	2	280/280

Table 4. Trajectory deviation results

		Actual Class		
		Class 1	Class 2	Class 3
Pred-	M1000	296/300	0	0
icted	M2000	0	291/300	16
class	M400	2	8	324/340

6 Conclusions

In this work, we have presented a framework for classification of Ballistic Missiles trajectory using ART-2 and HMM in real-time. Competitive learning of ART2 categorise the incoming samples into groups to be utilised by HMM for dynamic testing a new sample for its class and clubbing with classified group for training in the subsequent cycles. 6 DOF model generates trajectories of different ranges of ballistic missiles, which is utilized for testing and evaluation of ART2 and HMM. Making window of fix length allows HMM to integrate testing and training while taking care of variety of trajectories to be addressed by the same model. Further, we have established that the HMM is able to classify the new sample in less than 50 milliseconds which makes it to use for real-time applications with above 95% success rate on Pentium-4 processor.

References

1. Bashir, F.I., Khokhar, A.A., Schonfeld, D.: Object trajectory-based activity classification and recognition using hidden markov models. IEEE Transactions on Image Processing 16(7), 1912–1919 (2007)
2. Bishop, C.M.: Pattern reccognition and Machine Learning. Springer (2006)
3. Fausett, L.: Fundamentals of Neural Networks - Architectures, Algorithms and Applications. Prentice Hall (2008)
4. Fraser, A.M.: Hidden Markov Models and Dynamical Systems. Society for Industrial and Applied Mathematics (2008)
5. Carpenter, G.A., Grossberg, S.: Art 2: self-organization of stable category recognition codes for analog input patterns. Journal of Optics 26(23), 4919–4930 (1987)
6. Garcia, J., Poncha, O.P., Molina, J.M., de Miguel, G.: Trajectory classification based on machine-learning techniques over tracking data. In: 9th International Conference on Information Fusion, pp. 1–8 (2006)
7. Juang, B.H., Rabiner, L.R.: Hidden markov models for speech recognition. Technometrics 33(3), 251–272 (1991)
8. Mantle, P.J.: The Missile Defense Equation-Factors for decision making. American Institute of Aeronautics and Astronautics (2003)
9. Mlich, I.J., Chmelar, P.: Trajectory classification based on hidden markov models. In: 18th International Conference on Computer Graphics and Vision, pp. 101–105 (2008)
10. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of IEEE 77(2), 257–286 (1989)
11. Tait, P.: Introduction to Radar Target Recognition. The Institution of Engineering and Technology (2005)

Performance Analysis of Multiclass Common Spatial Patterns in Brain-Computer Interface

Soumyadip Chatterjee[1], Saugat Bhattacharyya[1], Amit Konar[1], D.N. Tibarewala[2], Anwesha Khasnobish[2], and R. Janarthanan[3]

[1] Department of Electronics and Telecommunication Engineering
Jadavpur University, Kolkata, India
[2] School of Bioscience and Engineering
Jadavpur University, Kolkata, India
[3] Department of Computer Science
TJS Engineering College, Chennai, India
{soumyadipc7,biomed.ju,anweshakhasno}@gmail.com,
saugatbhattacharyya@live.com,konaramit@yahoo.co.in,srmjana_73@yahoo.com

Abstract. Brain-Computer Interfacing (BCI) aims to assist, enhance, or repair human cognitive or sensory-motor functions. The classification of EEG signals plays a crucial role in BCI implementation. In this paper we have implemented a multi-class CSP Mutual Information Feature Selection (MIFS) algorithm to classify our EEG data for three class Motor Imagery BCI and have presented a comparative study of different classification algorithms including k-nearest neighbor (kNN) and Fuzzy kNN algorithm, linear discriminant analysis (LDA), Quadratic discriminant analysis (QDA), support vector machine (SVM), radial basis function (RBF) SVM and Naive Bayesian (NB) classifiers algorithms. It is observed that Fuzzy kNN and kNN algorithm provides the highest classification accuracy of 92.65% and 92.29% which surpasses the classification accuracy of the other algorithms.

Keywords: Brain-Computer Interfacing, Electroencephalography, Common Spatial Pattern, Mutual Information Features Selection, k-Nearest Neighbor, Fuzzy k-Nearest Neighbor, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Support Vector Machine, Nave-Bayesian.

1 Introduction

The main function of Brain-computer Interfacing (BCI) is to process and decode the brain signals and send the resulting commands to an external assistive device, thus implementing a real-time interface between the user and his environment. This interface may be a word processor, wheel chair or a prosthetic limb [1, 2]. In this technique, the subjects use their brain signals for communication and control of objects in their environment, thereby bypassing their impaired neuromuscular system [3, 4].

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 115–120, 2013.

Common Spatial Pattern (CSP) is instrumental in implementing extraction of intended activity from the neural recordings. CSP was first applied in BCI implementation in [5, 6]. BCI is generally limited to binary classification of data due to low information transfer rates. To enhance the information transfer rate one can move from binary to multiple classes. For this purpose, we have proposed a feature selection technique based on a simple multiclass CSP OVR [7] and Mutual Information Feature Selection (MIFS) [8] and have compared the performance of the proposed technique using seven different classification methods including k-Nearest Neighbor (kNN), Fuzzy kNN, Linear and Radial Basis Function (RBF-) Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Nave Bayesian (NB) classifiers [9 - 11] in differentiating the raw EEG data obtained into left/right hand and up movement. Our proposed framework also introduces a novel voting scheme to increase the classification accuracy.

The rest of the paper is structured as follows: Section 2 elucidates the proposed CSP-MIFS framework. The organization of the experimental data and data preprocessing is explained in section 3. Performance analysis of the classifiers is given in section 4. Section 5 concludes the paper.

2 Proposed Framework

The CSP algorithm was initially developed for binary classification of motor imagery. In this section we describe the binary CSP algorithm and extend its application for the multiclass case. The MIFS algorithm requires a user defined parameter k which denotes the number of features to be selected. It is based on the filter approach.

2.1 Proposed Approach of Multiclass CSP-MIFS

In this paper we consider the One-Versus-Rest (OVR) [7] approach to extend the CSP algorithm to multiple classes. As we are considering three classes, three CSP blocks have been employed. The input to the first CSP blocks will be the signals from class1 and a combination of class2 and class3 EEG data. Similarly, the input to the second classifier will be class2 and a combination of class3 and class1 EEG data and so on. Next, the CSP projection matrix for each of the 3 combinations will be computed and the spatially filtered matrix Z is created. The first 3 and last 3 rows are selected from Z and then they are subject to feature selection by MIFS algorithm [8].The spatially filtered and feature selected signals creates the feature vector to be fed to the classifiers. A comparative study of the classification accuracies of these algorithms are carried out. Finally, the classified data from each stage is processed by a voting mechanism which gives the final classes of the EEG data. This process is graphically shown in Fig. 1.

2.2 The Voting Mechanism

The input to the voting stage is the predicted classes of the data from the three classifiers. We denote the classes used in the first classifier as 1 and 23 where 1

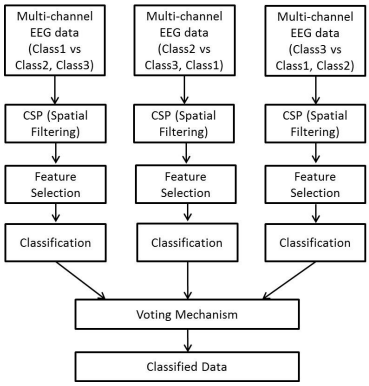


Fig. 1. Different stages of the proposed scheme for 3 classes

denotes the data is of class1 and 23 denotes the data is either of class2 or class3. Similarly the classes in the second classifier are denoted by 2 and 31 and for the third classifier as 3 and 12.c(x, y) is a function that computes the binary CSP between classes x and y and gives the predicted class. When there is an equal probability of a test data to belong to any one of two classes, then this function is called to resolve the matter. This happens in cases 2, 3 and 5. When the test data cannot be classified correctly to any of the classes, a random class is assigned to them. This is the case in 1 and 8. This is illustrated in Table 1.

Table 1. The Voting Mechanism

Sl. No.	Classifier 1	Classifier 2	Classifier 3	Class
1	1	2	3	rand(1,2,3)
2	1	2	12	c(1,2)
3	1	31	3	c(1,3)
4	1	31	12	1
5	23	2	3	c(2,3)
6	23	2	12	2
7	23	31	3	3
8	23	31	12	rand(1,2,3)

3 Data Analysis

All the experiments were conducted in our lab at Jadavpur University. 10 subjects (6 female and 4 male) performed the experiments in which they were instructed to imagine moving left (Class 1), right (Class 2) or forward (Class 3), according to the instructions displayed through a visual cue. The subjects performed the experiment in a single session, containing 120 trials each, i.e., 40 trials for each class.

Visual Cue. The visual cue is designed as follows: In the first 30 seconds of the session, a blank screen is displayed during which the baseline of the subject is measured followed by 60 trials of 6 seconds each. Each trial began with a fixation + for 1 second, which is an instruction to the subject to focus on the screen. Then a left/right/up arrow is displayed on the screen for 3 seconds as instruction to the subject. After 3 seconds, a blank screen would be displayed for 1.75-2.25 seconds to eliminate the cognitive effect of the current trial in the next one. The timing scheme of the visual cue is shown in Fig.2.

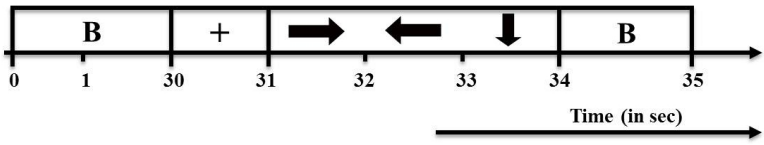


Fig. 2. Timing scheme of the visual cue displayed to the subjects

Experimental Setup. The EEG was recorded using an Emotiv Epoc system, which is a high resolution, multi-channel, wireless neuroheadset obtaining the EEG signals from the 14 electrode locations, based on the 10-20 electrode system. The electrode channels are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4. The sampling rate of the EEG system is 128 Hz.

Data Extraction and Pre-processing. Following the acquisition of the EEG signals, the raw EEG signal is band pass filtered using an IIR elliptical filter of order 6 between 8-24 Hz, as movement related signals are obtained from the 8-12 Hz mu- and 16-24 central beta band and to filter out any artifacts obtained during the recording. The filtered signal is further epoched into 1/16th of a second (i.e., 0.0625 seconds) and fed to the feature extraction and selection algorithm to form the feature vector.

Feature Extraction. Our proposed approach is applied to the epoched signal. First, the spatially filtered epoched signals are obtained and then fed to the MIFS feature selector to select the best features among the six rows. The final size of the features selected from each row is 4.

Classifiers. In this paper we have used both linear and non-linear classifiers which include k-Nearest Neighbor (kNN), Fuzzy kNN, Linear and Radial Basis Function (RBF-) Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Nave Bayesian (NB) classifiers [7-9].

4 Results and Discussion

The whole experiment is conducted in MATLAB version 7.9 environment. The specification of the system in which the experiment was conducted is as follows: Processor- Intel Core2Duo, 1.19 GHz, 3.2 GB RAM.

The total feature vector is partitioned into two different datasets, the training dataset and the testing dataset using k-fold cross-validation technique [10]. In our study, k is taken as 10. The feature vector is fed to the above classifiers and the classification accuracies are used as a parameter for performance analysis of the classifiers (as shown in Table 2). From Table 2, we observe that Fuzzy kNN gives the mean classification accuracy of 92.65% whereas kNN gives an accuracy of 92.29%.

Table 2. Average Classification Accuracy

Subjects	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Mean**	**Std**
kNN	96.15	95.31	92.43	90.17	93.59	90.72	89.56	94.57	85.11	95.29	**92.29**	**3.43**
Fuzzy kNN	96.88	95.35	92.40	90.32	93.65	90.71	89.66	94.70	86.77	96.01	**92.65**	**3.24**
SVM	72.10	73.52	75.21	76.22	73.89	74.20	75.37	72.66	75.12	72.98	**74.13**	**1.34**
RBF-SVM	85.12	84.56	82.35	82.01	83.23	81.36	81.02	83.14	78.85	84.93	**82.66**	**1.97**
LDA	65.87	67.23	70.52	71.41	70.25	71.17	72.07	68.03	74.21	66.10	**69.69**	**2.76**
QDA	81.27	80.78	79.54	77.16	80.16	76.98	76.85	80.51	72.29	80.92	**78.65**	**2.82**
NB	85.07	83.54	81.29	78.96	80.23	77.23	75.20	81.29	70.23	84.19	**79.72**	**4.54**

We have also employed Friedman Test [12] to statistically validate our results. The significance level is set at α =0.05. The null hypothesis here, states that all the algorithms are equivalent, so their ranks should be equal. We consider the mean classification accuracy, obtained from Table 2 as the basis of rank. Table 3 provides the ranking of each classifier algorithm.

Table 3. Ranking of the classifiers based on their average classification accuracy

Classifier j	**kNN**	**Fuzzy kNN**	**SVM**	**RBF-SVM**	**LDA**	**QDA**	**NB**
Rank	2	1	6	3	7	5	4

Now, from Table 3, we obtain r_j, $\chi_F^2 = 79.291$. Now, the χ_F^2 for our given dataset $> \chi^2_{7,0.05}= 14.067$. So, the null hypothesis, claiming that all the algorithms are equivalent, is wrong and, therefore, the performances of the algorithms are determined by their ranks only. It is clear from the table that the rank of Fuzzy kNN is 1, claiming Fuzzy kNN outperforms all the algorithms by Friedman Test.

5 Conclusion

The paper proposes a novel feature extraction and selection technique based on Multi-class Common Spatial Pattern and Mutual Information Feature Selection classification of multi-class problems. The resultant feature vector is fed to seven classifiers for a comparison on their performances. It is noted that Fuzzy kNN and kNN give the best results among all the classifiers and most of the classifiers give a result of more than 75%. Thus, our algorithm can be employed for further real time processing of multi-class problems. Further study in this direction will aim to optimize the feature selection, extraction and classification techniques to be implemented in real time application of Brain-Computer Interfacing.

Acknowledgments. I would like to thank University Grants Commission, India, University of Potential Excellence Programme (Phase II) in Cognitive Science, Jadavpur University and Council of Scientific and Industrial Research, India.

References

1. Lebedev, M.A., Nicolelis, M.A.: Brain-machine Interface: Past, Present and future. J. Trends Neurosci. 29(9), 536–546 (2006)
2. Vaughan, T.M., et al.: Brain-computer Interface Technology: A review of the second international meeting. IEEE Trans. Neural Sys. Rehab. Engg. 11(2), 94–109 (2003)
3. Daly, J.J., Wolpaw, J.R.: Brain-computer interfaces in neurological rehabilitation. Lancet Neurol. 7, 1032–1043 (2008)
4. Wolpaw, J.R., et al.: Brain-computer interfaces for Communication and Control. Clin. Neurophy. 113(6), 767–791 (2002)
5. Mller-Gerking, J., Pfurtscheller, G., Flyvbjerg, H.: Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin. Neurophy. 110(5), 787–798 (1999)
6. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehab. Engg. 8(4), 441–446 (2000)
7. Wu, W., Gao, X., Gao, S.: One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS), pp. 2387–2390. IEEE Press, Shanghai (2006)
8. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Net. 5(4), 537–550 (1994)
9. Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B.: A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces. J. Neural Eng. 4, R1–R13 (2007)
10. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Academic Press (2006)
11. Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. IEEE Trans. Systems, Man and Cybernetics SMC 15(4), 580–585 (1985)
12. Conover, W.J., Iman, R.L.: Rank transformations as a bridge between parametric and nonparametric statistics. The American Statistician 35(3), 124–129 (1981)

A Fuzzy Hybrid Framework for Offline Signature Verification

Geetha Ganapathi and R Nadarajan

P.S.G. College of Technology, Peelamedu,
Coimbatore, India
ngg@mca.psgtech.ac.in,
rn@amc.psgtech.ac.in

Abstract. Signatures are widely used means of personal verification. This paper presents a fuzzy hybrid framework based person-dependent off-line signature verification using fuzzy inference rules in image contrast enhancement, fuzzy rough reduction for feature selection and Simplified fuzzy ARTMAP for verification. Three sets of experimental studies are conducted on CEDAR benchmark dataset and the results reported are comparable to other techniques in terms of classification accuracy and time.

Keywords: Off-line signature verification, Simplified fuzzy ARTMAP, fuzzy inference rules, contrast intensification, fuzzy rough sets, feature selection.

1 Introduction

Signature is a widely used biometric for authentication of an individual or a document. A signature verification system aims to verify the identity of an individual based on the analysis of the signature. Signatures of some people vary substantially: even successive impressions of their signature are significantly different due to shape and relative position of the characteristic features. Hence, developing a robust signature verification system is a very challenging task. Signature verification is either on-line or off-line. Off-line verification lacks any form of dynamic information and has to rely on the features that are extracted from the static signature image and hence much more difficult to verify. Forgeries are random, simple or skilled. A skilled forgery is a close imitation of the original signature produced by a forger who has seen and practiced writing the genuine signature.

The complexity of signature verification lies in the variability of signing. The uncertainties related to the complexity of patterns and the lack of complete information about a signature motivated to develop a system for writer-dependent off-line signature verification for skilled forgeries. In this paper, the features of the signatures, the gray level intensities based on a gradient direction histogram are extracted and verified using Simplified fuzzy ARTMAP. As contrast intensification plays a major role in enhancing image quality, contrast intensification

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 121–127, 2013.

is done using fuzzy inference rules. As more features may introduce more measurement noise and reduce accuracy, relevant features are selected using fuzzy rough rules. All the algorithms are performed on the commonly used CEDAR [6] benchmark dataset and results reported.

The paper is organized is as follows. Section 2 discusses the related work. Section 3 discusses the proposed method. Section 4 elaborates on image enhancement, feature selection and verification. Section 5 reports on the experiments conducted and the results obtained. Finally, Section 6 concludes the work.

2 Related Work

Signature verification system is an active area of research. The Centre of Excellence for Document Analysis & Recognition (CEDAR) signature data set [6] is a commonly used data set for off-line signature verification. Schemes that used the CEDAR data set are referred here.

Kalera et al.[6] proposed an approach based on quasi-multi resolution technique using Gradient, Structural & Concavity (GSC) features for feature extraction and obtained 78% accuracy. Srihari et al.[14] developed a distance statistics method and acquired 78.1% accuracy. Chen et al.[1] proposed Zernike moments based features and distance similarity measure to get 83.6% accuracy. Chen et al.[2] presented a method based on graph matching, thin-plate spline mapping and word shape descriptors to get an accuracy of 90%. Larkins et al [10] introduced adaptive feature thresholding (AFT) combined with spatial pyramids [11] and equimass sampling grids [3] for feature extraction using gradient direction and achieved 90% accuracy. Kumar et al.[8] proposed a set of morphological features for extraction and multi-layer perceptron (MLP) and support vector machine (SVM) for classification and obtained an accuracy of 88.4%. Kumar et al.[9] proposed a set of features based on Surroundedness property of a signature image and based on MLP and SVM obtained an accuracy of 91.67%.

3 Proposed Fuzzy Hybrid Framework for Signature Verification

In Signature Verification, direct methods to generate features from image pixels or some form of transforms have been proposed in feature extraction and machine learning technique / similarity measure based approach for verification.

In this paper, we use grey level intensity and gradient direction histogram combined with spatial pyramids [11] and equimass sampling grids [3] for feature extraction as in [10]. We propose the following methodology:1) In preprocessing fuzzy rules based contrast enhancement is applied to increase the gray level intensity of the signature 2) gradient direction histogram combined with one level of equimass is used for feature extraction 3) Fuzzy Rough approach is applied for feature selection and 4) Simplified fuzzy ARTMAP is used for verification.

The feature extraction used in this paper depends on the gray level intensity of the signature image. A gray scale image possesses some ambiguity within the

pixels due to the possible multi-valued levels of grayness. Contrast enhancement improves the overall visibility of the image gray levels by transforming the dark pixels to appear darker and the light pixels to appear lighter. The advantage of the fuzzy methods compared to conventional image enhancement algorithm is their stability under nearly uncertain conditions on the one hand and the simple adjustment to reach a desired effect in the resulting image on the other hand; The fuzzy rules based approach is a powerful and universal method for many tasks in image processing [15]. Hassanien et al [4] have done a comparative study of 5 fuzzy based image enhancement algorithms on digital mammogram images and showed that fuzzy histogram hyperbolization method and fuzzy rule based system (FRBS) are comparatively better. Hence this paper uses FRBS for image enhancement with contrast intensification.

Feature Selection addresses the problem of selecting those input features that are most predictive of a given outcome. More features may introduce more measurement noise and, hence reduce accuracy. Hence an efficient and effective reduction is necessary. The success of rough set theory is due to finding a minimal representation of data without requiring additional information like thresholds. Rough Sets and Fuzzy sets can both tolerate inconsistency and uncertainty. Fuzzy sets are concerned with vagueness and rough sets are concerned with indiscernibility. Fuzzy-rough feature selection (FRFS), that employs fuzzy-rough sets to provide a means by which discrete or real-valued noisy data can be effectively reduced without the need for user-supplied information [5]. Rough set based approach provides a filter-based tool by which knowledge may be extracted from a domain in a concise way by using the granularity structure of the data. A reduced feature set is obtained by applying fuzzy Rough Reduct algorithm [5] to the set of features extracted from the signature image.

Traditional Neural Network algorithms like Radial Basis Function (RBF), MLP with BackPropagation have limitation on generalization. Fuzzy ARTMAP (FAM) neural network achieves a synthesis of fuzzy logic and adaptive resonance theory(ART). Kasuba's [7] Simplified FAM (SFAM) is a vast simplification over FAM. The fast learning capability of SFAM is thus utilized in classifying the signatures. The features extracted are given as inputs to a SFAM to determine if a signature is genuine or forged. False Acceptance Rate (FAR), a type II error that expresses the percentage of forgeries accepted as genuine and False Rejection Rate (FRR), a type I error that expresses the percentage of genuine rejected as forgery are used to compute the accuracy.

4 Verification of Signatures

The raw signature image undergoes preprocessing to reduce noise, feature extraction, feature subset selection and verification phases.

Preprocessing of Signatures: Preprocessing is a required step to modify the signature image to improve the representation of the extracted features. Each signature is binarized followed by Weiner noise reduction. Fig. 1.a represents the original image of a sample signature.

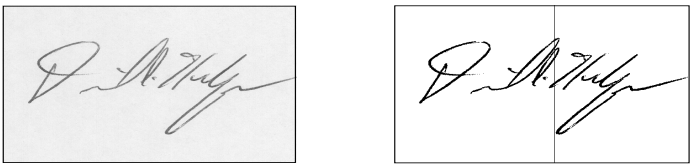


Fig. 1. Signature image : a. original b. one level of equimass

Feature Extraction: A feature vector is created for each signature based solely on the gradient direction of each pixel from across a signature. This direction θ of a pixel at (x,y) is found by $\theta = \tan^{-1}(\frac{G_x}{G_y})$ where G_x is the Sobel kernel for horizontal change and G_y for vertical change. The resulting direction value ranging between 0 to 2π radians, is split into 18 non-overlapping segments allowing a gradient direction histogram to be created from the count of each direction. Equimass is an adaptive grid based on the number of black pixels or Mass M of a signature, where the grid lines are found at the equimass divisions of the horizontal and vertical mass histogram. An effective approach as in [10] by combining the spatial pyramids [11] and the equimass sampling grids [3] were used. The features extracted as above are normalized. As each box (Fig.1.b) contains 18 segments, the signature image provides a first level of 36 features by equimass. Alg. 1 states the procedure for feature extraction and its time complexity would

Algorithm 1. Feature Extraction

1: **for** each image in a set **do**
2: Get the original image, apply Weiner filter to reduce noise
3: Apply fuzzification operator to the image
4: Apply constrast Intensification to the Fuzzified image
5: Apply defuzzification operator to the Intensified image
6: Divide the image into one level of equimass (2 boxes)
7: Get gradient direction histogram for the 18 segments (20deg) in each box
8: Normalize the data
9: **end for**

be O(n*w*h)where n would be the number of signatures per signer and w*h would be the size of the signature image. Steps 3, 4 & 5 are executed when fuzzy rule based contrast intensification is done.

Verification: Verification of Signatures is done by utilizing a SFAM [7]. SFAM contains two layers, the input and the output. The input to the network flows through the complement coder to the input layer. Weights from each of the output category nodes flow down to the input layer. The category layer holds the number of classes that the network has to learn. Vigilance parameter and match tracking mechanism are employed for network training. During training, the input data (36 features) are presented to the SFAM network together with their respective categories (genuine/forgery : 1/-1). During testing, the test input is

presented to the network. All the output nodes compute the activation functions with respect to the input. The winner is the node with the highest activation function and the input is classified to the category to which the winning output node belongs to. The training and testing phases of SFAM were adapted from [13]. The algorithms are applied for all 55 sets and the overall accuracy is computed. The time complexity for training algorithm would be O(n*e), where n is the number of training signatures and e is the number of epochs. The complexity for testing would be O(n), n being the number of test signatures.

Fuzzy Rule Based Contrast Intensification: The signature image is subjected to contrast intensification of gray levels using fuzzy inference rules as in [4]. A very simple inference system is formulated. The methodology is to first fuzzify the gray level, contrast intensify the fuzzy gray level and defuzzify the fuzzy value back to gray level values. Triangular membership functions are used for dark, gray and light fuzzy sets employed in the rules:

If pixel is dark then black
If pixel is gray then gray
If pixel is light then white

The fuzzy gray level is intensified by applying the INT operator [12] and is defuzzified back by inverse transformation. The complexity would be O(n*r*c) where n is the number of signatures and rxc is the size of the image.

Fuzzy Rough Feature Selection: The signature image based on gradient direction histogram and one level of equimass produce 36 features per signature. These features are fed into the fuzzy rough reduct algorithm [5] to select relevant features. The algorithm starts off with an empty set and adds in turn, one at a time, those attributes that result in the greatest increase in the rough set dependency metric, until this produces its maximum possible value for the dataset. The algorithm terminates when the addition of any remaining attribute does not increase the dependency. For n attributes, O(n*(n+1)/2) evaluations may be performed for the worst case data set. As the features are specific to a signer, we have a feature subset for each signer. A superset was created from these subsets which had 80% of features in common. The feature subset that got selected is {1, 2, 3, 4, 5, 14, 16, 17, 18, 19, 20, 21, 22, 23, 32, 33, 34, 36}. This set of 18 features, is a 50% reduction from our original feature set which consisted of features from 1 to 36.

5 Results and Discussion

The CEDAR signature dataset [6], a popular benchmark dataset used for off-line signature verification, consists of 55 signature sets, with each set being composed by one writer. Each set consists of 24 genuine and 24 skilled forgery signatures in a space measuring 2 x 2 inches, scanned at 300 dpi in 8-bit gray scale. In total, the dataset contains 1,320 genuine signatures and 1,320 forgeries. The experiment process was carried out on each signature set, where 14 each from genuine and forgery randomly chosen signatures were used as training set and the remaining 10 each in genuine and forgery were used for testing. The training

set keeps varying for each trial and 10 such trials were conducted in order to average the effect of possible outliers in the training set and also to obtain a reliable estimate of the accuracy. The evaluation was carried out on all the 55 sets. Table 1 compares the error rates and the accuracy against the existing methods, of which [8] and [9] are person independent. FRR keeps decreasing from FRBS contrast intensification to FRFS in the proposed methods implying that forgeries are being identified more accurately by contrast enhancement and feature selection. The vigilance parameter was 0.99 in SFAM. Table 2 lists the results when the training set is varied. All the experiments described above have been written in MATLAB running on Intel Core 2 Quad CPU @ 2.66 GHz PC with 2GB of RAM. Contrast intensification of Gray levels combined with spatial

Table 1. Results When Training set size is 16 (CEDAR data)

Method	Acc %	FRR	FAR
Existing			
Gradient Structural & Concavity [6]	78.50	22.45	19.50
Zernike Moments [1]	83.60	16.60	16.30
Graph Matching [2]	92.10	7.70	8.20
Adaptive Feature Thresholding [10]	90.44	1.86	10.96
Signature Morphology [8]	88.41	11.59	11.59
Surroundedness [9]	91.67	8.33	8.33
Proposed			
Only SFAM	88.95	6.27	15.83
FRBS Contrast + SFAM	92.85	1.71	12.60
FRBS + FRFS + SFAM	93.99	2.83	9.19

Table 2. Results for varying Training set size

Set Size	Acc %	FRR	FAR
10	92.79	3.30	11.13
11	92.97	3.85	10.21
12	93.47	3.57	9.48
13	93.65	3.45	9.25
14	93.99	2.83	9.19
15	94.13	2.69	9.04

pyramids [11] and equimass sampling grids [3] in feature extraction, and fuzzy rough sets in feature selection have helped to boost the classification accuracy.

6 Conclusion

Differentiating skilled forgeries from genuine is a very difficult task. In this paper, an offline signature verification system based on a fuzzy hybrid framework was proposed. The algorithms were tested on the benchmark CEDAR data set. The method used SFAM for verification and obtained an accuracy of 88.95%. FRBS

based contrast Intensification and enhancement improves the accuracy to 92.85% and FRFS further enhances the accuracy to 93.99%. With the system being related to skilled forgeries, it can be concluded that the fuzzy based framework augurs well in uncertain and ambiguous situations in terms of accuracy and time. Less number of features makes the system less computationally intense and this factor is more suitable for real-time systems.

References

1. Chen, S., Srihari, S.: Use of Exterior Contours and Shape Features in Off-line Signature Verification. In: Proc. of the 8th Internat. Conf. on Document Analysis and Recognition (ICDAR 2005), vol. 2, pp. 1280–1284 (2005)
2. Chen, S., Srihari, S.: A New Off-line Signature Verification Method based on Graph Matching. In: Proc. of the 18th Internat. Conf. on Pattern Recognition (ICPR 2006), vol. 2, pp. 869–872 (2006)
3. Favata, J., Srikantan, G.: A Multiple Feature/Resolution Approach to Hand printed Digit and Character Recognition. Int. J. Imaging Syst. Technol. 7(4), 304–311 (1996)
4. Hassanien, A.E., Badr, A.: A Comparative study on Digital Mammography Enhancement Algorithm based on Fuzzy Theory. Studies in Information and Control 12(1), 21–31 (2003)
5. Jensen, R., Shen, Q.: New Approaches to Fuzzy-Rough Feature Selection. IEEE Transactions on Fuzzy Systems 17(4), 824–838 (2009)
6. Kalera, M.K., Srihari, S., Xu, A.: Offline Signature Verification and Identification using distance statistics. Intern. J. Pattern Recognit. Artif. Intell. 18(7), 1339–1360 (2004)
7. Kasuba, T.: Simplified Fuzzy ARTMAP. AI Expert, 18–25 (November 1993)
8. Kumar, R., Kundu, L., Sharma, J.D., Chanda, B.: A writer-independent off-line signature verification system based on signature morphology. In: Proc. of the IITM 2010, pp. 261–265. ACM (2010)
9. Kumar, R., Sharma, J.D., Chanda, B.: Writer-independent off-line signature verification using surroundedness feature. Pattern Recognit. Lett. 33, 301–308 (2012)
10. Larkins, R., Mayo, M.: Adaptive Feature Thresholding for Off-line Signature Verification. In: 23rd Internat. Conf. in Image and Vision Computing New Zealand (IVCNZ 2008), pp. 1–6 (2008)
11. Lazebnik, S., Schmid, S., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit (CVPR), vol. 2, pp. 2169–2178 (2006)
12. Pal, S.K., King, R.A.: Image Enhancement using Smoothing with Fuzzy sets. IEEE Trans. Syst. Man Cybern. C Appl. Rev. 11(7), 494–501 (1981)
13. Rajasekaran, S., Vijayalakshmi Pai, G.: Neural Networks, Fuzzy Logic and Genetic Algorithm - Synthesis and Applications. Prentice Hall of India (2011)
14. Srihari, S., Xu, A., Kalera, M.K.: Learning Strategies and Classification Methods for Offline Signature Verification. In: Proc. of the 7th Internat. Workshop on Frontiers in Handwriting Recognition (IWHR), pp. 161–166 (2004)
15. Tizhoosh, H.R., Krell, G., Michaelis, B.: On Fuzzy Enhancement of Megavoltage images in Radiation Therapy. IEEE Conference of Fuzzy Systems, Barcelona, Spain 3, 1399–1404 (1997)

A Graph Based Approach to Multiview Clustering

Moumita Saha

Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur
moumitasaha@cse.iitkgp.ernet.in

Abstract. Rich and complex data sets prevalent in many applications can often be explored from multiple perspectives. Examples include clustering of multimedia, multilingual and heterogeneously linked data sets. Multiview clustering attempts to discover clusters from different views of the same data set. In this article, construction of subspace representation of the views and subsequently clustering the subspaces to produce multiview clusters is been proposed. The subspaces are obtained by a separate clustering procedure on the nearest neighbour graphs of the individual features. Three graph similarity measures are used for this clustering. Empirical results on three benchmark data sets shows that the proposed method provides superior performance in terms of classification accuracy using known class labels as compared to single view clustering of the entire data sets.

Keywords: Multiview clustering, graph measures, subspace clustering.

1 Introduction

Multi-view Clustering is an emerging topic in the field of data mining. In richly structured data the entities can be observed or modelled from various perspectives leading to multiple views or representations. Multi-view learning is an useful approach to effectively explore and exploit the information from heterogeneous data for the purpose of improving the learning performance. Multi-view algorithms deals with each view of the data independently and then merge the solutions to obtain a complete, robust pattern which is superior compared to its single-view representation.

In this paper, an approach for multiview clustering of data is proposed. A graph based methodology is adopted. Multiple clusterings of the graphs are obtained and used to represent the data from multiple views.

The paper is organised as follows. Review of some related work on multiview clustering is presented in section 2. The proposed graph based approach is delivered in section 3. Analysis and illustration of the experimental results is presented in section 4 and finally the paper is concluded in section 5.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 128–133, 2013.

2 Related Work

Multi-view clustering is becoming increasingly important in several application. Authors have proposed several approaches to address this problem. Among the earliest efforts, Bickel and Scheffer [1] propose partitioning and agglomerative, hierarchical multi-view clustering and apply them to text data. It is shown that multi-view versions of k-Means and EM outperform their respective single view as they optimize agreement between views. However, it shows negative result for agglomerative hierarchical clustering as the mixture components have smaller overlap when the views are concatenated.

In [2], both textual and visual contents of the retrieved images are considered and a multi-view clustering approach is proposed to re-rank the web retrieved result provided by text based search engine. Kim et al. [3] proposes a multiview clustering method for multilingual documents. The proposed approach is an incremental algorithm which first groups documents having the same patterns assigned by view-specific probabilistic latent semantic analysis (PLSA) models, and then the remaining un-clustered documents are assigned to the groups using a constrained PLSA model.

Greene and Cunningham [4] present an analysis of the research themes in a bibliographic literature network, based on the integration of both co-citation links and text similarity relationships between papers in the network. In [5], Bruno and Maillet investigated a late fusion approach for multi-view clustering based on the latent modelling of cluster-cluster relationships and it is shown to outperform an early-fusion approach based on multi-view feature correlation analysis.

Chaudhuri et al. [6] propose method for projection of higher dimensional data into its lower dimensional-subspace using multiple views of data by using Canonical Correlation Analysis (CCA). Results for mixtures of Gaussian and mixtures of log concave distributions are shown to be effective.

Several other approaches to multiview clustering are reported in literature. Many of them explore projected subspaces to obtain individual clustering views and finally combine them to provide different perspectives of the data. The graph based approach for finding appropriate single view subspace clusters for the data in multiview framework is described in the next section.

3 Graph Based Approach to Multi-view Clustering

Multi-view clustering is performed utilizing a graph representation of subspaces and hierarchical agglomerative clustering. A three stage approach is proposed for this. First, the data set is represented as graphs with nodes as data points and the link as feature wise similarity. Next various graph similarity measures is used to evaluate the similarity between the graphs and utilize them to cluster the graphs into different groups. Each of these clusters represents separate view or representation of the data. Finally, for each of the view, clustering on the corresponding projected subspaces is applied in order to obtain final clusters of instances for each of the views. The schema of the approach is shown in figure 1.

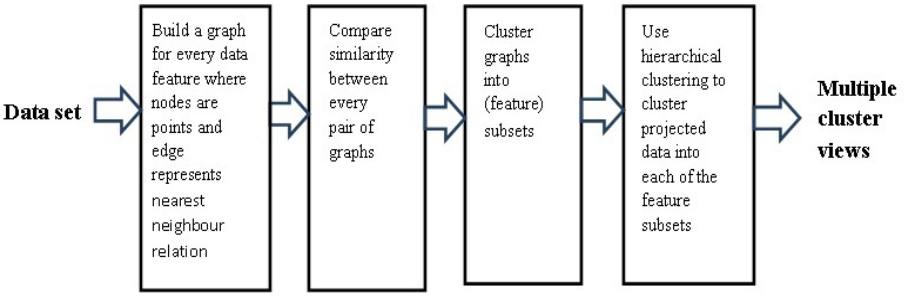


Fig. 1. Block diagram of graph theoretic multiview clustering

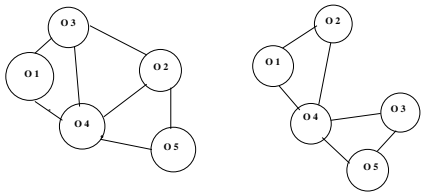
3.1 Feature Graph Construction

The data set is represented as graph where nodes are the data points and edges exist between nearest neighbour nodes. d number of graphs is build for a d-dimensional dataset representing each of the features of the data set. As a pre-processing step all the duplicate values of the data points are removed and data points are arranged in sorted order in order to obtain the nearest neighbour fast.

Illustration of representation of a toy data set is shown in table 1.

Table 1. Illustration of an example graph representation

Point	F1	F2
O1	2	200
O2	7	190
O3	5	150
O4	6	180
O5	12	160
O6	2	200



3.2 View Extraction

The subsequent step is view extraction of the data. Different views are extracted by assembling the similar features together. Agglomerative clustering is used to cluster the features into different groups which will represent the views. Similarity measures considered are the following.

Cumulative Degree Centrality (CDC) : The degree centrality of a node v, for a given graph $G := (V, E)$ with $\mid V \mid$ vertices and $\mid E \mid$ edges, is defined as $C_D(\mathrm{v}) = \deg(\mathrm{v})$.

If n is the number of nodes of the graph G then cumulative degree centrality is evaluated as in equation 1.

$$Cumulative\ degree\ centrality\ of\ G\ =\ \Sigma_{i=1}^{n}\ deg(v_i) \tag{1}$$

Edges Count of the Graph (EC) : Edges are the connections between the vertices of the graph. This set is denoted by $E(G)$. If $G := (V, E)$ represents a graph, individual edges are pairs $\{u, v\}$ where u and v are vertices in V. The total number of edges of each feature graph is considered as a measure of the corresponding graphs.

Similar Edge Count (SEC) : The third approach taken is evaluating the number of similar edges between two graphs.

If $G := (V, E)$ and $F := (V', E')$ are two graphs, then two edges $(u, v) \in E$ and $(u', v') \in E'$ are said to be similar if their end vertices are same, that is either

$$u = u' \quad \text{and} \quad v = v' \qquad \text{or} \qquad u = v' \quad \text{and} \quad v = u'.$$

Thus, each pair of graph is compared on the basis of number of similar edges between the graphs.

Three more measure are evaluated which are combination of the above measures. They are combination of degree centrality and number of similar edges (CDC + SEC), combination of degree centrality and number of edges (CDC + EC), combination of degree centrality, number of edges and number of similar edges (CDC + EC + SEC) between the graphs.

The approach considered is as follows:

1. The above mentioned measures are evaluated for each of the feature graph.
2. This measures are taken as the similarity measure for executing agglomerative clustering to obtain the cluster of features.
3. Dendograms obtained from clustering results are analysed and the features are grouped into different clusters. Each cluster of features represents a view of the data. Three clusters, i.e. three views of the data based on best accuracy performance are considered.

3.3 Clustering the Views

The final step of the algorithm is clustering the instances according to the different views obtained to get a better analysis of the group of instances with respect to the views. The approach followed is depicted below.

1. The data set is projected according to the features obtained in each view.
2. The reduced projected view of data is clustered utilizing complete linkage agglomerative clustering. The number of cluster is set by user. If class label is available it is set as the number of classes.

The accuracy of clustering is evaluated in terms of classification accuracy as follows.
(a) Evaluate the count of each class label in each group of instances.
(b) Calculate the maximum count and assign the particular group with that maximum count class label.
(c) Repeat the process until all the group is assigned a class label, keeping in mind no two group can be assigned the same class label.

The above process is repeated for every view of the data and this multi-view approach is compared to its single view counterpart.

4 Result and Analysis

The proposed graph-based approach to multiview clustering is applied on three datasets - SPECTF Heart dataset, Ozone Layer dataset and SECOM dataset. The description of dataset is shown in table 2.

Table 2. Datasets description

Dataset	No. instances	No. features	No. classes
SPECTF Heart	267	44	2
Ozone	2537	72	2
SECOM	1561	590	2

The result is depicted in two major steps. First the outcome of views extraction is evaluated and then the final accuracy of clustering of instances according to multiple views are established.

Table 3 shows as an example the resulting views of the SPECTF Heart data obtained from clustering of the feature graph taking into consideration different similarity measures. The accuracy of each views of three data sets are depicted in tables 4, 5, 6. Single view clustering is performed by taking all the features of the data to cluster the data points and the result is compared to multiview clustering outcome. It is observed that the accuracy improves in case of multiview clustering as compared to its single-view counterpart. This fact is depicted in table 7

Table 3. Multiple views of SPECTF Heart

Similarity Measures	view 1	view 2	view 3
CDC	1, 8	2-6, 9-44	7
EC	31	1-24, 27-40	25-26, 41-44
SEC	5	2-4,6-44	1
CDC + SEC	1, 8	2-6, 9-44	7
CDC + EC	31	1-24, 27-40	25-26, 41-44
CDC + EC + SEC	31	1-24, 27-40	25-26, 41-44

Table 4. Accuracy (%) of different views for SPECTF Heart

Similarity Measures	view 1	view 2	view 3
CDC	74.44	72.22	72.96
EC	77.03	72.22	73.33
SEC	75.18	64.81	62.92
CDC + SEC	74.44	72.22	72.96
CDC + EC	77.03	72.22	73.33
CDC + EC + SEC	77.03	75.92	73.70

Table 5. Accuracy (%) of different views for Ozone

Similarity Measures	view 1	view 2	view 3
CDC	97.32	94.17	94.07
EC	85.03	84.28	85.03
SEC	78.08	94.10	62.17
CDC + SEC	94.17	97.32	94.07
CDC + EC	86.14	60.07	94.10
CDC + EC + SEC	85.03	84.28	94.07

Table 6. Accuracy (%) of different views for SECOM

Similarity Measures	view 1	view 2	view 3
CDC	86.29	87.82	82.70
EC	81.74	93.01	93.08
SEC	93.27	91.15	93.27
CDC + SEC	74.95	87.82	82.70
CDC + EC	93.01	75.78	86.80
CDC + EC + SEC	93.08	71.49	85.07

Table 7. Accuracies (%) of single vs multiview clustering

Data set	Single view clustering	Multi view clustering	Best view of multi view clustering	Corresponding similarity measure
SPECTF Heart	73.70	77.03	view 1	EC & CDC + EC & CDC + EC + SEC
Ozone	94.10	97.32	view 2	CDC + SEC
SECOM	91.15	93.27	view 1 & view 3	SEC

5 Conclusion

A graph based approach is presented for multiview clustering. Nearest neighbour graphs of data point projected to individual features are first constructed using three different similarity measures between these graphs, they are clustered to feature subspaces. The subspaces correspond to individual views. Hierarchical agglomerative clustering is performed on each of these views to obtain multiview clusters. The multiview clusters are evaluated in terms of their classification accuracy and is found to provide superior performance than single view clustering. In future, use of other graph similarity measures and clustering algorithms may be explored to improve performance.

References

1. Bickel, S., Scheffer, T.: Multi-View Clustering. In: Proceeding of 4th IEEE International Conference on Data Mining, pp. 19–26 (2004)
2. Chi, M., Zhang, P., Zhao, Y., Feng, R., Xue, X.: Web Image Re-Ranking with Multi-view Clustering. In: Proceeding of 18th Int. World Wide Web Conference, pp. 1189–1190. ACM (2009)
3. Kim, Y.-M., Reza Amini, M., Goutte, C., Gallinari, P.: Multi-View Clustering of Multilingual Documents. In: Proceeding of 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 821–822 (2010)
4. Greene, D., Cunningham, P.: Multi-View Clustering for Mining Heterogeneous Social Network Data. In: Proceeding of Workshop on Information Retrieval over Social Networks, 31st European Conference on Information Retrieval, ECIR 2009 (2009)
5. Bruno, E., Maillet, S.M.: Multiview Clustering: A Late Fusion Approach using Latent Models. Proceedings of 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2009, 736–737 (2009)
6. Chaudhuri, K., Kakade, S.M., Livescu, K., Sridharan, K.: Multi-View Clustering via Canonical Correlation Analysis. In: Proceedings of 26th Annual International Conference on Machine Learning, ICML 2009, pp. 129–136 (2009)

Classification of Fricatives Using Novel Modulation Spectrogram Based Features

Kewal D. Malde, Anshu Chittora, and Hemant A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology,
Gandhinagar, Gujarat, India
{kewal_dhiraj,anshu_chittora,hemant_patil}@daiict.ac.in

Abstract. In this paper, we propose the use of a novel feature set, *i.e., modulation spectrogram* for *fricative classification*. Modulation spectrogram gives *2*-dimensional (*i.e.*, *2-D*) feature vector for each phoneme. Higher Order Singular Value Decomposition (HOSVD) is used to reduce the size of large dimensional feature vector obtained by modulation spectrogram. These features are then used to classify the fricatives in five broad classes on the basis of place of articulation (*viz., labiodental*, *dental*, *alveolar*, *post-alveolar* and *glottal*). Four-fold cross-validation experiments have been conducted on TIMIT database. Our experimental results show *89.09 %* and *87.51 %* accuracies for recognition of place of articulation of fricatives and phoneme-level fricative classification, respectively, using *3-nearest neighbor classifier*.

Keywords: Fricative classification, modulation spectrogram, HOSVD, place of articulation, acoustic frequency and modulation frequency.

1 Introduction

The main purpose of this paper is to use modulation spectrogram-based features to distinguish between various fricatives sounds extracted manually from TIMIT database. Consonants are classified in three broad classes based on the manner of articulation, *viz.*, stops, affricates and fricatives. Fricatives class is the largest class of consonants in TIMIT database. In total, there are *10* fricatives divided in five broad classes based on *place* of articulation, *viz.*, labiodental (*/f/*, */v/*), dental (*/th/*, */dh/*), alveolar (*/s/*, */z/*), post-alveolar (*/sh/*, */zh/*) and glottal (*/hh/*, */hv/*). Fricatives can also be classified as *unvoiced* and *voiced* on the basis of voicing activity, *i.e.*, relaxed or vibrating vocal folds. From all the examples of the fricatives mentioned above in pairs, the former in each pair falls under unvoiced fricatives while later falls under voiced fricatives [1-4].

Fricatives are generated by a narrow constriction in the vocal tract, giving rise to a steady *frication noise*. The degree of constriction also plays an important role in spectral characteristics (as secondary effect with vocal tract being the primary one). Place of articulation and the spectral characteristics are influenced on the basis of location of constriction by the tongue (at the back, center, or front of the oral tract, as well as at the teeth or lips) [1].

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 134–139, 2013.

Prior work on fricative classification on TIMIT database is based on the knowledge-based, acoustic-phonetic features where Seneff's auditory model is used as front-end processing. An accuracy of *87 %* was obtained for the overall classification of fricatives (on training data set) [5-7].

In this paper, we introduce the use of *2-D* representation of modulation spectrogram based for classification of fricatives. Modulation spectrogram-based feature set integrates the concept of sensory perception with signal processing methodology to achieve a significant improvement in the representation and coding of acoustic signals [8-9]. Recently, modulation spectrogram has been used in voice quality classification, pathology classification and speech recognition [10].

2 Modulation Spectrogram

Modulation spectrogram is the visual representation of the spectrum of the combination of *acoustic* and *modulation* frequencies in a speech segment. The spectrogram is square of the magnitude of short-time Fourier transform (STFT) of speech signal, representing the energy distribution of the signal *w.r.t. time* and *acoustic* frequency parameters. Modulation spectrogram is the spectrum of magnitude spectra of speech segment, representing energy distribution of the signal w.r.t. *acoustic* and *modulation* frequency parameters. Each phoneme has different combination of *acoustic* and *modulation* frequencies (depending on the constriction formed by the tongue). This motivated us to use modulation spectrogram for fricative classification. Modulation spectrogram is given by [10],

$$X_m(k) = \sum_{n=-\infty}^{\infty} h(mM - n)x(n)W_{I_A}^{kn}, \tag{1}$$

$$X_l(k,i) = \sum_{m=-\infty}^{\infty} g(lL - m) \mid X_m(k) \mid W_{I_M}^{im}, \tag{2}$$

where $k = 0, 1, ..., (I_A\text{-}1)$ is *acoustic* frequency, $W_{I_A} = e^{-j(2\pi/I_A)}$, $W_{I_M} = e^{-j(2\pi/I_M)}$, $i = 0, 1, ..., (I_M\text{-}1)$ is *modulation* frequency; n is the sample index in time-domain, m is the frame index in equation *(1)* and l is the frame index in equation *(2)*; $h(n)$ and $g(m)$ are the analysis window for acoustic and modulation frequency with hop sizes M and L, respectively. Here, $|X_l(k,i)|^2$ represents the modulation spectrogram of signal $x(n)$.

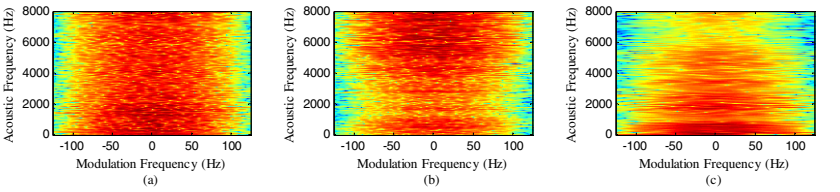


Fig. 1. Modulation Spectrogram: (a) labiodental */f/*, (b) alveolar */s/*, (c) glottal */hh/*

From some of the examples of modulation spectrogram shown in Fig. 1, we can observe that phonemes from different classes have their energy distributed across different regions of acoustic frequency (f_a) and modulation frequency (f_m) (as the mode of constriction formed by the tongue is different for each phoneme); hence, it can serve as a potential *cue* for classification of fricative sounds.

3 Higher Order Singular Value Decomposition (HOSVD)

Higher Order Singular Value Decomposition (HOSVD) is applied to *3-D* feature set. HOSVD theorem is used for dimension reduction of a *3-D* tensor as described in [11]. First of all, modulation spectrogram which is a *2-D* feature vector is obtained for each phoneme segment. Then a *3-D* tensor is formed by *stacking 2-D* modulation spectrogram feature vector (*i.e.*, $B \in \mathbb{R}^{I_A \times I_M}$ by considering only positive modulation frequencies taking symmetric nature of modulation spectrogram into consideration) for all phoneme samples (say I_S) under each class. Thus, the tensor $A \in \mathbb{R}^{I_A \times I_M \times I_S}$ can be represented in HOSVD form as:

$$A = S \times_1 U_A \times_2 U_M \times_3 U_S, \tag{3}$$

where S is the core tensor with same dimension as A. $U_A \in \mathbb{R}^{I_A \times I_A}$, $U_M \in \mathbb{R}^{I_M \times I_M}$ and $U_S \in \mathbb{R}^{I_S \times I_S}$ are the *unitary* matrices of the corresponding subspaces of I_A, I_M and I_S. The matrices U_A and U_M can be obtained by *unfolding* tensor A. U_A and U_M are obtained from SVD representation of the unfolded matrices $A_A \in \mathbb{R}^{I_M \times I_A I_S}$ and $A_M \in \mathbb{R}^{I_A \times I_M I_S}$. Then, $\hat{U}_A \in \mathbb{R}^{I_A \times R_A}$ and $\hat{U}_M \in \mathbb{R}^{I_M \times R_M}$ are obtained from these unitary matrices U_A and U_M, by retaining only first R_A and R_M vectors, respectively. Now, using these obtained $\hat{U}_A$ and $\hat{U}_M$, we get the reduced tensor as,

$$\hat{A} = S \times_1 \hat{U}_A \times_2 \hat{U}_M \times_3 U_S. \tag{4}$$

And finally un-stacking the reduced tensor $\hat{A}$, reduced feature vector $K \in \mathbb{R}^{R_A \times R_M}$ (per phoneme) is obtained as,

$$K = B \times_1 \hat{U}_A^T \times_2 \hat{U}_M^T = \hat{U}_A^T \cdot B \cdot \hat{U}_M. \tag{5}$$

4 Experimental Setup

4.1 Database Used

TIMIT database contains a total of *6300* sentences, *10* sentences spoken by each of *630* speakers from *8* major dialect regions of the United States. In addition to a speech waveform (*i.e.*, **.wav*) file, TIMIT corpus includes three associated transcription files (*viz.*, **.txt*, **.wrd*, **.phn*). With the help of phonetic transcription (*i.e.*, **.phn*) file we

obtained fricative segments. Out of all obtained fricative segments, we randomly selected *1000* samples of each phoneme (from both training and testing dataset) under *fricative* class. In our work, we have considered all *10 fricatives.*

4.2 Feature Extraction

First of all, find the modulation spectrogram for each phoneme using COG (center of gravity) method [12]. In modulation spectrogram, Y-axis represents *acoustic* frequency which ranges from 0 to $F_s/2$ (here, $F_s = 16$ *kHz*) (equally spaced in *54* bins) and X-axis represents *modulation* frequency which ranges from *-100* Hz to *+100*Hz (equally spaced bins). The modulation frequency dimension varies with respect to phoneme class and phoneme duration. For making the classification task easier, f_m dimension is fixed to *140* by *zero padding*. Since the modulation spectrogram is symmetric around $f_m = 0$, only positive modulation frequencies is considered for analysis. Using the *symmetry* property of modulation spectrogram the feature size is reduced from 54*x*140 to 54*x*71. To further reduce the feature dimension, HOSVD theorem is applied which gives a feature vector of dimension $f_a \times f_m$.

5 Experimental Results

Four-fold cross-validation experiments have been conducted on *56* different combinations of f_a and f_m. We have considered following values of f_a and f_m , *viz.*, f_a = [3 4 5 6 7 8 9] and f_m = [3 4 5 6 7 8 9 12]. In order to quote statistical significance of our experimental results, complete phoneme set is *randomly* divided into training and testing datasets (as per training-testing ratio), for every experiment.

Results of *4*-fold cross-validation of recognition of place of articulation of fricatives and phoneme-level fricative classification with feature vector dimension of 3×3 and *75:25 %* training-testing ratios are shown in Table 1 and Table 2, respectively. Fig. 2 shows the effect on classification accuracy due to change in acoustic (f_a) and modulation (f_m) frequencies. From Fig. 2, we can see that initially accuracy increases as the value f_a and f_m is increased (as the phoneme are more closely related in spectral characteristics, we need relatively higher dimension of feature vector for better classification of data), however, after certain amount of increase in the values of f_a and f_m , any further increase in the value decreases the classification accuracy (as it introduces more amount of redundancies). It can also be observed that for lower value of f_m (i.e., $f_m = 3$ and $f_m = 4$) accuracy is *almost constant*. We can also infer that change in the value of f_m affects the accuracy more significantly as compared to the change in the value of f_a. Best classification accuracy is obtained for feature of dimension 3×8 . However, we have selected the optimum feature vector dimension as 3×3 since the improvement in the accuracy is approximately *1 %* while the dimension is reduced by a factor of approximately *2.67*. The reduction in feature vector dimension is advantageous in reducing computational *complexity* and computation *time*.

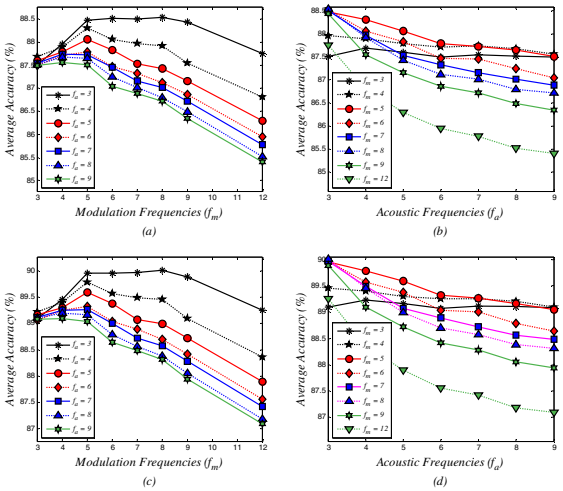


Fig. 2. Average accuracy (in %) for fricative consonant classification at phoneme-level: Effect on classification accuracy due to change in (a) modulation frequency (f_m), (b) acoustic frequency (f_a); Average accuracy (in %) for recognition of place of articulation of fricatives: Effect on recognition accuracy due to change in (c) modulation frequency (f_m), (d) acoustic frequency (f_a). In all sub-figures, units of acoustic and modulation frequencies are in bins.

Table 1. Confusion matrix for accuracy (in %) for recognition of place of articulation of fricatives. Average accuracy is *89.09 %*.

			Detected				
			Labiodental	Dental	Alveolar	Post-alveolar	Glottal
			/f/, /v/	*/th/, /dh/*	*/s/, /z/*	*/sh/, /zh/*	*/hh/, /hv/*
Actual	Labiodental	*/f/, /v/*	**88.22**	3.43	2.97	1.97	3.41
	Dental	*/th/, /dh/*	3.45	**89.85**	2.34	1.41	2.94
	Alveolar	*/s/, /z/*	2.92	2.34	**88.27**	3.44	3.03
	Post-alveolar	*/sh/, /zh/*	2.01	1.37	3.48	**90.81**	2.33
	Glottal	*/hh/, /hv/*	3.43	2.97	2.98	2.32	**88.30**

Table 2. Confusion matrix for phoneme-level fricative classification accuracy (in %). Average accuracy is *87.51 %*.

		Detected									
		/f/	*/v/*	*/th/*	*/dh/*	*/s/*	*/z/*	*/sh/*	*/zh/*	*/hh/*	*/hv/*
Actual	*/f/*	**85.89**	1.64	2.71	1.13	1.67	1.75	1.50	0.57	1.91	1.23
	/v/	1.61	**87.30**	1.65	1.37	1.22	1.29	1.29	0.58	1.78	1.90
	/th/	2.73	1.68	**85.70**	1.30	1.63	1.63	1.36	0.44	2.02	1.51
	/dh/	1.08	1.42	1.32	**91.37**	0.70	0.73	0.75	0.27	1.26	1.10
	/s/	1.63	1.24	1.55	0.71	**86.21**	2.10	2.66	1.01	1.44	1.45
	/z/	1.67	1.30	1.64	0.79	2.08	**86.15**	2.28	0.94	1.48	1.68
	/sh/	1.53	1.37	1.37	0.70	2.71	2.24	**85.39**	1.24	1.49	1.95
	/zh/	0.58	0.55	0.43	0.23	0.98	1.03	1.21	**93.77**	0.52	0.71
	/hh/	1.93	1.78	2.11	1.24	1.39	1.50	1.44	0.54	**86.48**	1.59
	/hv/	1.28	1.87	1.43	1.16	1.41	1.67	1.96	0.70	1.73	**86.80**

6 Summary and Conclusions

In this paper, we have proposed the use of modulation spectrogram-based features using a simple classifier, *i.e.*, *3*-nearest neighbor classifier, for fricative classification. Modulation spectrogram gives good classification accuracy for recognition of place of articulation of fricatives and phoneme-level fricative classification. Feature vector dimension of 3×3 after applying HOSVD theorem is considered as optimum. Modulation frequency parameter plays more important role as compared to acoustic frequency parameter, as it affects the classification accuracy more significantly. One of the limitations of present work is that, we have worked on manually segmented fricative segments. In future, we would like to extend this work on continuous speech.

Acknowledgments. The authors would like to thank Department of Electronics and Information Technology (DeitY), New Delhi (India) and authorities of DA-IICT, Gandhinagar for their support to carry out this research work. We would also like to thank Mr. Maulik Madhavi for his support in the preparation of the paper.

References

1. Quatieri, T.F.: Discrete-time Speech Signal Processing: Principles and Practice. Prentice Hall Press, Upper Saddle River (2004)
2. Web Source, `http://www.langsci.ucl.ac.uk/ipa/IPA_chart_C2005.pdf` (last accessed on 30th April, 2013)
3. Garofolo, J.S.: Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. National Institute of Standards and Technology (NIST), Gaithersburgh, MD (1988)
4. Scanlon, P., Ellis, D., Reilly, R.: Using Broad Phonetic Group Experts for Improved Speech Recognition. IEEE Trans. on Audio, Speech and Language Proc. 15, 803–812 (2007)
5. Ali, A.M.A., Spiegel, J.V., Mueller, P.: Acoustic-phonetic features for automatic classification of fricatives. J. Accoust. Soc. of America 109(5), 2217–2235 (2001)
6. Ali, A.M.A., Spiegel, J.V., Muller, P.: An acoustic-phonetic feature-based system for the automatic recognition of fricative consonents. In: IEEE Proc. on Int. Conf. on Acoustics, Speech and Signal Processing, vol. 2, pp. 961–964 (1998)
7. Seneff, S.: A Joint Synchrony/ Mean Rate Model of Auditory Speech Processing. J. Phonetics 16, 55–76 (1988)
8. Atlas, L., Shamma, A.S.: Joint acoustic and modulation frequency. EURASIP J. on Applied signal Proccessing 7, 668–675 (2003)
9. Greenberg, S., Kingsbury, B.: The modulation spectrogram: In pursuit of an invariant representation of speech. In: IEEE Proc. on Int. Conf. on Acoust., Speech, Signal Process., Munich, Germany, vol. 3, pp. 1647–1650 (1997)
10. Markaki, M., Stylianou, Y.: Voice pathology detection and discrimination based on modulation spectral features. IEEE Trans. on Audio, Speech, and Language Proc. 19(7), 1938–1948 (2011)
11. Lathauwer, L.D., Moor, B.D., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. 21(4), 1253–1278 (2000)
12. Modulation Toolbox, `http://www.ee.washington.edu/research/isdl/projects/modulationtoolbox` (last accessed on 30th April 2013)

Speaker Recognition Using Sparse Representation via Superimposed Features

Yashesh Gaur, Maulik C. Madhavi, and Hemant A. Patil

Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, Gujarat, India
{yashesh_gaur,madhavi_maulik,hemant_patil}@daiict.ac.in

Abstract. In this paper, we demonstrate the effectiveness of *superimposed features* for the purpose of template matching-based speaker recognition using sparse representations. The principle behind our hypothesis is, if the test template approximately lies in the *linear span* of the training templates of the genuine class, then so does any linear combination of test templates. In this paper, we introduce the notion of *superimposed features* for the first time. Using our initial trials on the TIMIT database, we have shown that *superimposed features* can result in reducing the complexity cost by *80* % with a very minor decrease in identification rate by *0.67* % and a minor increase in EER by *0.85* %.

Keywords: Superimposed features, sparse representations, orthogonal matching pursuit, template matching, speaker recognition.

1 Introduction

Speaker recognition is the task to recognize a person from his or her voice, with the help of machines. Depending on the way feature matching is done, the systems can be classified into template matching systems and probabilistic modelling systems [1]. Probabilistic modelling systems involve modelling feature vectors with *probability density functions* (pdf). The probability of a test utterance, given the speaker model, is evaluated to get the *confidence scores* [1]-[2]. The template matching techniques, on the other hand, do not involve any probabilistic measures. The features from the test utterances are considered as some variation of the training features [1]. The template matching-based techniques are usually faster as no probabilistic modelling is required prior to matching. A sparse representation for the purpose of pattern classification has been used in [3]-[5]. A sparse representation model for the speaker recognition was used in [6]. This technique was probabilistic in nature as they used Gaussian Mixture Model (GMM) mean super vectors to model speaker characteristics. Sparse representations were invoked after the speaker characteristics were modelled using GMM. Recently, sparse representations were used using template matching technique in [7]. The benefit of this kind of technique is the reduction in complexity as sparse representations can directly be used on the features, without any prior

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 140–147, 2013.

modelling. The work also uses Orthogonal Matching Pursuit (OMP)[8] for the sparse recovery of the weight vector.

However, due to large amount of features in a test or training utterance, this approach sometimes becomes computationally ineffective as sparse recovery has to be performed for each and every test feature vector. In this paper, we propose a new method for speaker recognition using sparse representations, via *superimposed features*, which substantially reduces the computational complexity of such a system. The main assumption behind our approach is, if the testing template lies in the linear span of the training templates, then so does any linear combination of the testing templates. Thus, superimposed features can be considered as test features. This simple but powerful insight can help us in making real-time template matching systems for speaker recognition which are computationally much less expensive, with a very little degradation in performance.

The rest of the paper is organized as follows: The sparse representation framework, sparse recovery using OMP, evaluation of confidence scores and superimposed features are discussed in Section *2*. In Section *3* and Section *4*, we discuss the experimental setup and results, respectively. Finally, Section *5* concludes the work along with future research directions.

2 Sparse Representation Framework and Superimposed Features

2.1 Sparse Representation Framework

Suppose that each speaker is being evaluated against K speakers and each speaker has a set of N training features, which are m-dimensional. Let us define

$$A_k = [\mathbf{a}_{k1}, \mathbf{a}_{k2}, .., \mathbf{a}_{kN}] \ \in \ \mathbb{R}^{m \times N}, \tag{1}$$

as a $m \times N$ matrix of all the training features of k^{th} speaker concatenated. Here, $\mathbf{a}_{kn}$ represents the feature vector extracted from n^{th} frame of the k^{th} speaker. A universal dictionary can thus be created by concatenating such matrices from all the K speakers. That is,

$$A = [A_1, A_2, .., A_K] \ \in \ \mathbb{R}^{m \times K \cdot N}. \tag{2}$$

Now, consider a feature vector $\mathbf{y}$, extracted from a test utterance of some speaker. One can express this feature vector as an approximate linear combination of columns of matrix $\mathbf{A}$ as

$$\mathbf{y} \approx \sum_{k=1}^{K} \sum_{n=1}^{N} x_{kn} \mathbf{a}_{kn}, \tag{3}$$

where x_{kn} is the weight associated with the column a_{kn}. This can be written in a more compact form as

$$\mathbf{y} = A\mathbf{x} + \mathbf{n}, \tag{4}$$

where the vector $\mathbf{x}$ contains the weights corresponding to the columns of matrix A, $\mathbf{n}$ is the *noise vector* which accounts for the approximation in (3) and also

the noise present in the measurements. If $\mathbf{y}$ belongs to a particular class, say the k^{th} class, then it can be said that it will approximately lie in the linear span of the training vectors of the k^{th} class [6],[7]. In other words, the test vector of the k^{th} class can be represented as a linear combination of the training vector of the k^{th} class. In that case, the weight vector should exhibit high sparsity since most of the weights corresponding to vectors of other classes will be zero and only vectors from the k^{th} class will have non-zero weights [6],[7].

To find the weights corresponding to the columns of A, we need to solve the system of linear equations $\mathbf{y} = A\mathbf{x}$. Since $m \ll K \cdot N$, this is an underdetermined set of equations and there are infinitely many solutions to this system. However, we know that the weight vector needs to be highly sparse, therefore, we need to choose the sparsest solution among these infinitely many solutions. In other words,

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = A\mathbf{x}. \tag{5}$$

This optimization problem is an NP hard problem [9]. However, several greedy algorithms like matching pursuit [8], Orthogonal Matching Pursuit (OMP) [8], etc. have been proposed to solve this system. In this work, we will be using OMP for the sparse recovery of the weight vector $\mathbf{x}$.

2.2 Sparse Recovery Using Orthogonal Matching Pursuit (OMP)

A brief description of the OMP algorithm is given in Box B.1. The algorithm with its performance guarantees and details can be found in [8].

Box B.1. OMP Algorithm

Initialize: $\hat{\boldsymbol{x}}_0 = 0$, $\boldsymbol{r}_0 = \boldsymbol{y}$ *and* $\Lambda_0 = \Phi$
for $i = 1$; $i := i + 1$ *until stopping criterion is met,*
Do
Form signal estimate from residual: $\boldsymbol{g}_i \longleftarrow A^T \boldsymbol{r}_{i-1}$
Add largest residual entry to support: $\Lambda_i \longleftarrow \Lambda_{i-1} \cup supp(max(\boldsymbol{g}_i))$
Update signal estimate using least squares: $\hat{\boldsymbol{x}}|_{\Lambda_i} = A^{\dagger}_{\Lambda_i} \boldsymbol{y}|_{\Lambda_i^c \longleftarrow 0}$
Update measurement residual: $\boldsymbol{r}_i = \boldsymbol{y} - A\hat{\boldsymbol{x}}_i$
End for

This algorithm takes the test vector $\mathbf{y}$, the matrix A and returns the approximated weight vector $\hat{\mathbf{x}}$. In each iteration, it selects the column in A which has the highest correlation with the residual, adds the column index to the set Λ, gets the signal estimate by applying least squares to the columns of A indexed in the set Λ, and updates the residual for the next iteration. The stopping criterion used here is,

$$\|\mathbf{r}_i\|_2 < \lambda \|\mathbf{y}\|_2. \tag{6}$$

Therefore, whenever the norm of the residue goes below a certain fraction of $\|\mathbf{y}\|_2$, the algorithm stops. Heuristically, we found that $\lambda = 0.1$ best suits the requirements of this sparse recovery. It should be noted that this value will differ

for different scenarios. The value of λ must be chosen carefully, since a high value will not be able to capture the contributions of enough vectors, whereas a very small value may result into OMP running unnecessarily for a longer duration.

2.3 Confidence Scores

Suppose, we are given a frame extracted from a testing utterance. We can find the weight vector $\hat{\mathbf{x}}$, using OMP algorithm as described in the previous section. This vector will be highly sparse as only a few vectors from the genuine class will contribute to this test frame. The contribution of each speaker to the test frame can be calculated using class-based residual [3],[6],[7]. The class-based residual for k^{th} speaker can be computed by retaining the weights corresponding to the k^{th} class, and putting all other weights to zero. This can be done by defining a function $\delta_k(\hat{\mathbf{x}})$as:

$$\delta_k(\hat{\mathbf{x}}) = [0 \dots 0|0 \dots 0| \dots |\hat{x}_{k1} \dots \hat{x}_{kN}| \dots |0 \dots 0]\,. \tag{7}$$

The normalized class-based residual error can be computed as:

$$\mathbf{r}_k(y) = \frac{\|\mathbf{y} - A\delta_k(\hat{\mathbf{x}})\|_2}{\|\mathbf{y}\|_2}. \tag{8}$$

The normalized class based residual error always lies between *0* and *1*. A residual error closer to *0* indicates a close match. The confidence score for k^{th} speaker for test frame $\mathbf{y}$ can be computed as:

$$c_k(\mathbf{y}) = exp\left(-r_k(\mathbf{y})\right). \tag{9}$$

This is the confidence score of the k^{th} speaker, for a single frame $\mathbf{y}$, extracted from the test utterance. Suppose there are Z frames in the testing utterance. Confidence scores are calculated for all the Z frames and a mean confidence score is generated for the k^{th} speaker. That is,

$$C_k = \frac{1}{Z}\sum_{i=1}^{Z} c_k\left(\mathbf{y}_i\right). \tag{10}$$

This process can be repeated for all the K speakers to get speaker-specific confidence scores for the given test utterance.

2.4 Superimposed Features

If we have a frame from a test utterance belonging to the k^{th} class, it approximately lies in the linear span of the training vectors of that class [6],[7]. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be two frames which come from a testing utterance of k^{th} speaker. Both $\mathbf{y}_1$ and $\mathbf{y}_2$ can be approximately written as a linear combination of training vectors of the k^{th} class.

$$\mathbf{y}_1 \approx \alpha_{11}\mathbf{a}_{k1} + \alpha_{12}\mathbf{a}_{k2} + \dots + \alpha_{1N}\mathbf{a}_{kN}, \tag{11}$$

$$\mathbf{y}_2 \approx \alpha_{21}\mathbf{a}_{k1} + \alpha_{22}\mathbf{a}_{k2} + \cdots + \alpha_{2N}\mathbf{a}_{kN}, \tag{12}$$

$$\mathbf{y}_1 + \mathbf{y}_2 \approx (\alpha_{11} + \alpha_{21})\,\mathbf{a}_{k1} + (\alpha_{12} + \alpha_{22})\,\mathbf{a}_{k2} + \cdots + (\alpha_{1N} + \alpha_{2N})\,\mathbf{a}_{kN}. \tag{13}$$

It follows that, any combination of $\mathbf{y}_1$ and $\mathbf{y}_2$ should also approximately lie in the linear span of training vectors of the k^{th} class. Let us call the combination of $\mathbf{y}_1$ and $\mathbf{y}_2$ as the '*superimposed vector*'. If this '*superimposed vector*' is used to calculate the weights, the weight vector $\hat{\mathbf{x}}$ will still be *sparse*, as ideally, only the vectors from the genuine class should contribute to the test frame (13). This weight vector after class-based residual should give identical results, as in the case of $\mathbf{y}_1$ or $\mathbf{y}_2$ alone. The entire argument can be extended for superposition of more than two frames as well. Let us call any such combination of multiple test frames as the '*superimposed vector*' and define '*level of superposition*' as the number of feature vectors that are superimposed.

One can see that, using a superimposed feature, with level of superposition *'L'*, implies that we are clubbing L features together and feeding it to OMP. Consequently, the number of times OMP is invoked is reduced by L times. This in turn reduces the computational complexity of the system, and can help us in making such systems real time. It should be noted that, though the weight vector is still sparse, its sparsity will change after superimposed test frames are given. OMP can be similarly used for the sparse recovery of the weight vector, however, the stopping criterion on the residual may need to be changed. For our experiments, we have used the same threshold for stopping OMP.

3 Experimental Setup

3.1 Speech Corpus

The proposed approach is applied on all the *630* speakers in the TIMIT corpus. In the corpus, each speaker has *10* waveforms, which are divided into '*sx*', '*sa*' and '*si*' sections, containing *5*, *2* and *3* files per speaker, respectively [10]. The features, from the waveforms in the section '*sx*' and '*sa*' are taken for training purpose, whereas the features from the waveforms in '*si*' section are taken for testing purposes. Thus, there are *630* test trials. Each test is evaluated against *20* speakers, which are randomly selected. However, the correct speaker is always present in the *20* claimants.

3.2 Feature Extraction

Mel-Frequency Cepstral Coefficients (MFCC) [11] were used as features to represent speaker-specific characteristics. The speech signal is pre-emphasized and then divided into frames of *20* ms with an overlap of *10* ms. Each frame is first multiplied by a Hamming window and then *26* Mel-scaled triangular filters are used to extract a *19*-dimensional MFCC vector. This vector is appended by

delta cepstral coefficients which are computed over a span of ± 2 frames, thus producing a *38*-dimensional feature vector. Voice Activity Detector (VAD) is also used to discard frames with low energy. Cepstral Mean Subtraction (CMS) and normalization were also performed on the feature vectors to cancel the channel effects.

3.3 Performance Measures

The performance of the proposed technique can be tested using various parameters. We have used Equal Error Rate (EER) [12], % Identification rate (%ID) and computational complexity as the standard measures for performance analysis.

3.3.1. % Equal Error Rate (EER)

The confidence score obtained by evaluating test speaker with genuine speaker is called as *true* (genuine) score and all other confidence scores are called *impostor* scores. For speaker verification, an operating threshold score can be set for making decisions. In the decision making process, errors are encountered and the system may face *2* types of errors, *viz.*, false acceptance (false alarm) and false rejection (miss detection). For one particular operating threshold, both the errors are equal and the corresponding error value (false alarm or miss detection) is called as Equal Error Rate (EER). % EER is very useful evaluation metric in the speaker verification task [12].

3.3.2. % Identification Rate

Let N_C be the total number of correct hits (i.e., number of correctly identified speakers) using the algorithm and N_T be total number of test trials (i.e., total number of speakers). % Identification (%ID) rate is defined as following.

$$\%ID = \frac{N_C}{N_T} \times 100. \tag{14}$$

3.3.3. Computational Complexity

Computational complexity is one of the most important constraints when real-time systems are deployed. The major complexity step in the mentioned scheme is the OMP step. Therefore, the computational complexity can either be measured by number of OMP invocations or the average time it takes to process one query. All experiments were conducted on *Intel Xeon (R), CPU E5-2420 @ 1.90 GHz* machine.

4 Experimental Results

Table 1 gives a detailed description of the effect of various levels of superposition on the performance of system.

It is evident from Table 1 that with increase in level of superposition, there is a very minor deterioration in the values of % EER and % ID. Comparing to the

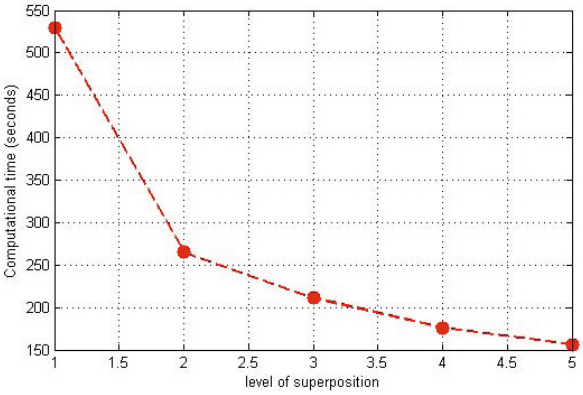


Fig. 1. Average execution time per query, in seconds, for different level of super position

Table 1. % EER, % ID and execution time for different levels of superposition on TIMIT database

Level of superposition	**% EER**	**% ID**	**Execution time (sec)**
1	0.45	99.84	530
2	0.45	99.84	265
3	0.7	99.67	211
4	0.97	99.51	176
5	1.3	99.17	156

baseline results, superposition level *5* results in a minor EER increase of *0.85* % and the % ID decreases by just *0.67* %. On the other hand, the numbers of OMP invocations are reduced to one-fifth of the original value. Consequently, the execution time per query is also reduced, which is also depicted by Figure 1. Hence, one can significantly reduce the computational cost of the system, with a little deterioration in EER and identification rate. This experiment can be conducted for higher level of superposition as well. It is natural that the % ID rate and EER performance will be deteriorated as level of superposition increases. This is because, the approximation depicted in (13) grows more and more inaccurate with increasing level of superposition. Therefore, a trade-off exists between computational complexity and EER and % ID rate performance. If one goes on increasing the level of superposition, one will reach a point when the EER and % ID performance is no longer acceptable. That should be the optimal level of superposition for the system. The optimal level of superposition will differ from system to system and has to be experimentally determined. In this paper, we have considered $\%ID > 99$ % as acceptable performance. Thus, we have limited ourselves to superposition level of *5*.

5 Summary and Conclusions

In this paper, we have proposed a new technique for speaker recognition, using sparse representations via superimposed features, which greatly reduce the complexity cost with a minor deterioration in EER and % ID rate performance. However, the proposed method does not take into account the increased sparsity of the weight vector when superimposed vectors are fed to OMP. Presently, we are running OMP using the same stopping criterion every time. OMP might need to run a few more iterations, as the sparsity has been decreased. Work can be done which adaptively determines the number of iterations needed for a particular level of superposition. The robustness of this scheme can also be checked against noise with varying SNR.

Acknowledgments. The authors would like to thank Department of Electronics and Information Technology (DeiTy), Govt. of India for their support to carry out this work.

References

1. Campbell Jr., J.P.: Speaker recognition: a tutorial. Proc. of the IEEE 85(9), 1437–1462 (1997)
2. Hazen, T., et al.: Multi-modal Face and Speaker Identification on a Handheld Device. In: Proc. Wkshp. Multimodal User Authentication, pp. 120–132 (2003)
3. Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(2), 210–227 (2009)
4. Pillai, J.K., et al.: Secure and Robust Iris Recognition Using Random Projections and Sparse Representations. IEEE Trans. on Pattern Analysis and Machine Intelligence 33(9), 1877–1893 (2011)
5. Yang, A.Y., et al.: Distributed recognition of human actions using wearable motion sensor networks. J. of Ambient Intelligence and Smart Environments 1(2), 103–115 (2009)
6. Naseem, I., Togneri, R., Bennamoun, M.: Sparse Representation for Speaker Identification. In: 20th Int. Conf. on Pattern Reco. (ICPR), pp. 4460–4463 (2010)
7. Boominathan, V., Sri Rama Murty, K.: Speaker recognition via sparse representations using orthogonal matching pursuit. In: Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP), pp. 4381–4384 (2012)
8. Elad, M.: Sparse and Redundant Representations. Springer, New York (2009)
9. Zucker, S.W., Leclerc, Y.G., Mohammed, J.L.: Continuous Relaxation and Local Maxima Selection: Conditions for Equivalence. IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI 3(2), 117–127 (1981)
10. Garofolo, J.S.: Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database. National Institute of Standards and Technology (NIST), Gaithersburgh, MD (1988)
11. Davis, S., Mermelstein, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech and Signal Process. 28(4), 357–366 (1980)
12. Martin, A., et al.: The DET Curve in Assessment of Detection Task Performance. In: Proc. Eurospeech 1997, vol. 4, pp. 1899–1903 (1997)

Kernel Based Rough-Fuzzy C-Means

Rohan Bhargava and Balakrushna Tripathy

School of Computing Science and Engineering, VIT University
Vellore 632014, Tamil Nadu, India
{rb.bhargava,tripathybk}@gmail.com

Abstract. Data clustering has found its usefulness in various fields. Algorithms are mostly developed using euclidean distance. But it has several drawbacks which maybe rectified by using kernel distance formula. In this paper, we propose a kernel based rough-fuzzy C-Means (KRFCM) algorithm and use modified version of the performance indexes (DB and D) obtained by replacing the distance function with kernel function. We provide a comparative analysis of RFCM with KRFCM by computing their DB and D index values. The analysis is based upon both numerical as well as image datasets. The results establish that the proposed algotihtm outperforms the existing one.

Keywords: Clustering, Kernel, DB Index, Dunn Index, Rough-Fuzzy C-Means.

1 Introduction

Cluster is a collection of data elements that are similar to each other but dissimilar to elements in other clusters. Cluster analysis is a key tool in the field of data analysis. Clustering techniques have their use in areas like analysis of statistical data, pattern recognition, image analysis, information retrieval, bioinformatics and data mining. Clustering algorithms partition data into a certain number of groups or so called clusters. There is no set of predefined rules to determine the correctness of clustering. Hence many variations can be made to any single algorithm to develop a new algorithm. An iterative technique of partitioning a dataset into K-clusters was introduced by MacQueen in 1967 [1]. Applying this concepts of fuzzy sets Ruspini [2] first proposed the fuzzy clustering algorithm, which was later modified and generalized by Dunn [3] and Bezdek [4] respectively. Similarly using the concept of rough sets P. Lingras proposed the rough k-means clustering algorithm[5]. Further developments led to the proposal of rough set based kernel k-means algorithm by Zhou et al. [6] and Tripathy et al. [7]-[8].

Distance between objects can be calculated in many ways, the euclidean distance based clustering is easy to implement and hence most commonly used. It has two drawbacks, firstly the final results are dependent on the initial centers and secondly it can only find linearly separable cluster. Kernel based clustering helps in rectifying the second problem as it produces nonlinear separating hyper surfaces among clusters [9] . Kernel functions are used to transform the data in

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 148–155, 2013.

the image plane into a feature plane of higher dimension known as kernel space. Nonlinear mapping functions used transforms the nonlinear seperation problem in the image plane into a linear serperation problem in kernel space facilitating clustering in feature space. Mercer's theorem [10] can be used to calculate the distance between the pixel feature values in kernel space without knowing the transformation function.

It was pointed out by Dubois and Prade [11] that rough and fuzzy sets complement each other. In fact the hybrid model of rough fuzzy and fuzzy rough sets provide a better model for representing imperfect data. In fuzzy set theory we have definite formulae for the computation of membership values. Thus the hybrid algorithms takes care of both features by providing membership values to elements as well as modeling vagueness in data through the boundary concept. The concepts of lower and upper approximations in rough set deals with uncertainity and vagueness, whereas the concept of membership function in fuzzy set helps in enhancing and evaluating overlapping clusters.

In this paper we implement and further modify the Rough-Fuzzy C-Means given by Maji et al. [12] to propose a new hybrid kernel based algorithm. We show the comparison between the two using numeric datasets and image datasets. The paper contains 5 sections. Section 2 provides the basic information about the euclidean and kernel distance functions. Section 3 gives a detailed explanation on the proposed kernel based Rough-Fuzzy C-Means algorithm. Section 4 is where the evaluation results are discussed. Finally the paper is concluded in section 5.

2 Types of Distance Functions

Euclidean Distance. The euclidean distance $d(x, y)$ between any two objects x and y in any n-dimensional plane can be found using

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}. \tag{1}$$

where, $x_1, x_2 \ldots x_n$ and $y_1, y_2 \ldots y_n$ are attributes of x and y respectively.

Kernel Distance. If x is an object then $\phi(x)$ is the transformation of x in high dimensional feature space where the inner product space is defined by $K(x, y) = \langle \phi(x), \phi(y) \rangle$. In this paper we use the Gaussian kernel function.

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right). \tag{2}$$

Where, $\sigma^2 = \sum_{k=1}^{N} \|x_k - \overline{x}\|^2 / N$ with $\overline{x} = \sum_{k=1}^{N} x_k / N$.

N is total number of data objects [9]. According to Phillips et al. [13] kernel distance function $D(x, y)$ in the generalized form is $D(x, y) = K(x, x) + K(y, y) - 2K(x, y)$ and on applying the property of similarity (i.e., $K(x, x) = 1$) it can be further reduced to (3).

$$D(x, y) = 2(1 - K(x, y)). \tag{3}$$

3 Kernel Based Rough-Fuzzy C-Means (K-RFCM)

Rough Fuzzy C-Means was proposed by P. Maji et al.[12] and S. Mitra et al. [14]; it combines the concepts of rough set theory and fuzzy set theory. The concepts of lower and upper approximations in rough set deals with uncertainty, vagueness and incompleteness whereas the concept of membership function in fuzzy set helps in enhancing and evaluating overlapping clusters. We follow the same concept and replace all euclidean distance functions with kernel distance function given in (3). According to rough set theory if $x_j \in \underline{B}U_i$ then object x_j is contained completely in cluster U_i and if $x_j \in BN(U_i)$ then object x_j belongs to cluster U_i and also belongs to the boundary of another cluster. Hence, according to fuzzy set theory the objects in boundary of clusters will have different membership values for the concerned clusters. Hence, membership values of objects in lower approximation are $\mu_{ij} = 1$ while for those in boundary region are the same as that in FCM. The steps followed in this algorithm are given below

1. Assign initial means v_i for c clusters.
2. Compute μ_{ik} using

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (\frac{D_{ik}}{D_{jk}})^{\frac{2}{m-1}}} \tag{4}$$

3. Let μ_{ik} and μ_{jk} be the maximum and next to maximum membership values of object x_k to cluster centroids v_i and v_j.
   If $\mu_{ik} - \mu_{jk} < \delta$ then
   $x_k \in \overline{B}U_i$ and $x_k \in \overline{B}U_j$ and x_k cannot be a member of any lower approximation.
   Else $x_k \in \underline{B}U_i$.
   where, delta is given by

$$\delta = \frac{1}{N} \sum_{i=1}^{N} (\mu_{ik} - \mu_{jk}) \tag{5}$$

4. Calculate new cluster means by using

$$V_i = \begin{cases} w_{low} \times A + w_{up} \times B & if\ |\underline{B}U_i| \neq \phi \text{ and } |BN(U_i)| \neq \phi; \\ B & if\ |\underline{B}U_i| = \phi \text{ and } |BN(U_i)| \neq \phi; \\ A & ELSE \end{cases} \tag{6}$$

$$\text{where, } A = \frac{\sum_{x_k \in \underline{B}U_i} x_k}{|\underline{B}U_i|} \text{ and } B = \frac{\sum_{x_k \in \overline{B}U_i - \underline{B}U_i} \mu_{ik}^m x_k}{\sum_{x_k \in \overline{B}U_i - \underline{B}U_i} \mu_{ik}^m}$$

5. Repeat from step 2 until termination condition is met of until there is no more assignment of objects

DB and Dunn Index. The Davis-Bouldin (DB) and Dunn (D) indexes [15] are two of the basic performance indexes. They help in evaluating the efficiency of clustering. The results are depend upon the number of clusters required. The DB index is defined as the ratio of sum of within-cluster distance to between-cluster distance. It is formulated as

$$DB = \frac{1}{c}\sum_{i=1}^{c} max_{i\neq j}\left\{\frac{S(U_i)+S(U_j)}{d(U_i,U_j)}\right\} \qquad \text{for } 1<i,j<c \tag{7}$$

A good clustering procedure should give value of DB index as low as possible. Similar to the DB index the D index is used for the identification of clusters that are compact and separated. It is computed as

$$Dunn = min_i\left\{min_{j\neq i}\left\{\frac{d(U_i,U_j)}{max_l S(U_l)}\right\}\right\} \qquad \text{for } 1<i,j,l<c \tag{8}$$

Greater value for the D index proves to be more efficient The within cluster distance denoted be $S(U_i)$ is given as

$$S(U_i) = \begin{cases} w_{low}\times C + w_{up}\times D & if\ |\underline{B}U_i| \neq \phi \text{ and } |BN(U_i)| \neq \phi; \\ D & if\ |\underline{B}U_i| = \phi \text{ and } |BN(U_i)| \neq \phi; \\ C & ELSE \end{cases} \tag{9}$$

$$\text{where, } C = \frac{\sum_{x_k\in\underline{B}U_i} {D_{ik}}^2}{|\underline{B}U_i|} \text{ and } D = \frac{\sum_{x_k\in\overline{B}U_i-\underline{B}U_i} \mu_{ik}^m {D_{ik}}^2}{\sum_{x_k\in\overline{B}U_i-\underline{B}U_i} \mu_{ik}^m}$$

4 Evaluation

The evaluation has been done in 2 parts. Firstly using a few real datasets and then on image datasets. We compare the results from RFCM algorithm and the proposed K-RFCM algorithm. Each datset is compared on the basis of DB and Dunn index.

4.1 Numerical Dataset: Iris and Soybean

Table 1 shows the cluster centers that are formed after applying each algorithm on the iris dataset consisting of 50 elements having 4 attributes each. Evaluation has been performed for c = 2,3 and 4. The initial centers are taken to be the first c elements of the dataset. We see that the results of both algorithms have very minute difference. Further analysis has been done for $c = 2$ in Table 2, it shows which all elements lie in the lower and boundary regions of each cluster. Again esults are approximately the same except for some values that have been highlighted in the table. These values are responsible for the minute difference observed in Table1.

Table 3 and 4 provides a comparison of the methods RFCM and KRFCM based upon the computations of DB and D index on the iris data set and the

Table 1. Cluster Center Values on Iris Dataset ($c = 2$,3 & 4)

No. of Clusters		RFCM	K-RFCM
2	Center 1	5.1639; 4.3754; 3.3974; 2.6100	5.1531; 4.3497; 3.3873; 2.6047
	Center 2	4.8498; 3.9619; 3.1373; 2.9359	4.8373; 3.9638; 3.1324; 2.3930
3	Center 1	5.1571; 4.3643; 3.3929; 2.6036	5.1308; 4.3538; 3.3897; 2.6019
	Center 2	4.8927; 3.9783; 3.1430; 2.3980	4.8935; 3.9788; 3.1431; 2.3981
	Center 3	4.5961; 3.9278; 3.0708; 2.3600	4.5984; 3.9296; 3.0719; 2.3608
4	Center 1	5.1571; 4.3643; 3.3929; 2.6036	5.1308; 4.3538; 3.3897; 2.6019
	Center 2	4.9150; 3.9850; 3.1400; 2.4059	4.9146; 3.9990; 3.1573; 2.4089
	Center 3	4.7134; 3.9620; 3.0776; 2.3589	4.7157; 3.9645; 3.0791; 2.3602
	Center 4	4.6128; 3.8615; 3.0740; 2.3561	4.6528; 3.9027; 3.1164; 2.3880

Table 2. Lower and Boundary Elements for Iris Dataset ($c = 2$)

Cluster Center	Lower		Boundary	
	RFCM	K-RFCM	RFCM	K-RFCM
Center 1	0, 4, 7, 10, **16**, 17, 19, 21, 27, 28, 36, 39, 40, 46, 48	0, 4, 7, 10, 17, 19, 21, **26**, 27, 28, **31**, 36, 39, 40, 46, 48	**2**, 5, 6, 8, 11, 13, 14, 15, 18, 20, 22, 23, 24, **26**, 29, **31**, 32, 33, 35, 38, 41, 42, 43, 44, 47, 49	5, 6, 8, 11, 13, 14, 15, **16**, 18, 20, 22, 23, 24, 29, 32, 33, 35, 38, 41, 42, 43, 44, 47, 49
Center 2	1, 3, 9, 12, 25, 30, 34, 37, 45	1, **2**, 3, 9, 12, 25, 30, 34, 37, 45	**2**, 5, 6, 8, 11, 13, 14, 15, 18, 20, 22, 23, 24, **26**, 29, **31**, 32, 33, 35, 38, 41, 42, 43, 44, 47, 49	5, 6, 8, 11, 13, 14, 15, **16**, 18, 20, 22, 23, 24, 29, 32, 33, 35, 38, 41, 42, 43, 44, 47, 49

Table 3. DB and Dunn Indexes for Iris Dataset

No. of Clusters	DB Index		Dunn Index	
	RFCM	K-RFCM	RFCM	K-RFCM
2	115.8219	15.6873	0.0159	0.1275
3	157.5086	20.7628	0.0093	0.0743
4	309.0443	84.4577	0.0042	0.01365

soybean dataset. In Table 3 it is clear that the values for DB index in K-RFCM are far lower than those of RFCM and the values for Dunn index are larger in the former algorithm. Hence stating that while results are similar the performance of K-RCCM is better than RFCM. Looking over at Table 4 the soybean data set consists of 37 elements having 35 attributes, the DB index values are lower for K-RFCM and Dunn index values are also low for the same. Though larger values for Dunn index were expected but we predict the low values are due to the large number of attributes involved in the dataset.

Table 4. DB and Dunn Indexes for Soybean Dataset

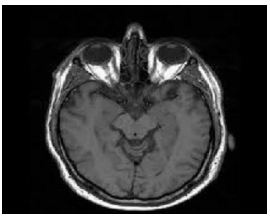
No. of Clusters	DB Index		Dunn Index	
	RFCM	K-RFCM	RFCM	K-RFCM
2	624.7929	119.7455	0.0029	0.0162
3	1074.2871	116.8911	9.4638	0.0096
4	1737.3794	192.9677	6.4533	0.0059

Table 5. DB and Dunn Indexes for Image datasets

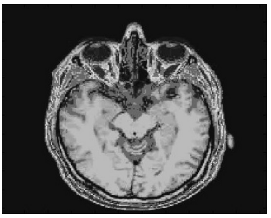
No. of Clusters	DB Index		Dunn Index	
	RFCM	K-RFCM	RFCM	K-RFCM
Brain	4.0469	0.0594	0.1060	10.3673
Cell	0.0561	negligible	11.4630	very large
Iris	4.7198	0.1643	0.1643	5.4695
Penny	9.2726	0.2369	0.0916	4.9417

4.2 Image Dataset

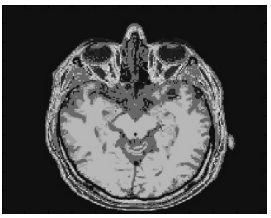
We have processed a number of images using both the algorithms to obtain a resultant image and as well as the DB and Dunn index values. Table 5 gives us the overview of the DB and Dunn index values obtained for each image. In all cases we observe that we achieve our desired results. There is a significant drop in DB index and significant increase in Dunn index. Comparing each image indivdually. Fig. 1c is sharp and has more clarity in outlines and finite details than that of Fig 1b. The difference between Fig 2b. and Fig 2c. is highly evident. Cell can be correctly identified in Fig. 2c. Finally, there is no noticeable change seen in Fig 3b. and Fig 3c.



1a : Original

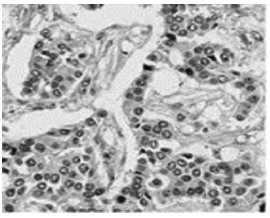


1b : RFCM Version

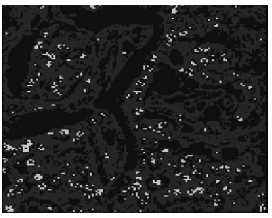


1c : K-RFCM Version

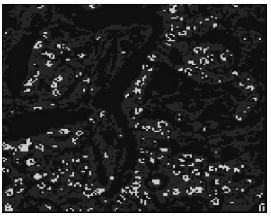
Fig. 1. Brain Image



2a : Original

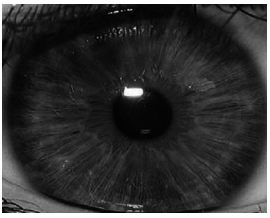


2b : RFCM Version

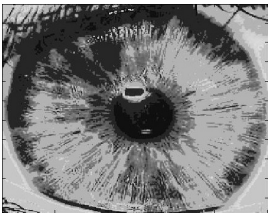


2c : K-RFCM Version

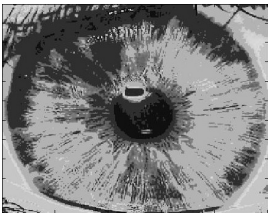
Fig. 2. Cell Image



3a : Original



3b : RFCM Version



3c : K-RFCM Version

Fig. 3. Iris Image

5 Conclusion

This paper focuses on improving the performance of the existing RFCM algorithm by using kernel function instead of euclidean distance. Hence, developing a new hybrid kernel based algorithm. Also, two of the most widely used performance indexes have been modified using kernel distance function for the evaluation of kernel based algorithms. Comparison between RFCM and proposed K-RFCM has been done on a wide variety of datasets to obtain favourable results. From the obtained results we can conclude that the proposed algorithm clearly outperforms the existing algorithm on the basis of performance and yields equivalent or better outputs in image dataset. The DB and D index introduced in this paper can also be applied on kernel based algorithms using rough sets and fuzzy sets individually to compare their performances.

References

1. MacQueen, J.B.: Some Methods for classification and Analysis of Multivariate Observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297. University of California Press (1967)
2. Ruspini, E.H.: A new approach to clustering. Information and Control 15(1), 22–32 (1969)
3. Dunn, J.C.: A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, 32–57 (1973)
4. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers (1981)
5. Lingras, P., West, J.: Interval set clustering of web users with rough k-means. Journal of Intelligent Information Systems 23(1), 5–16 (2004)
6. Zhou, T., Zhang, Y., Lu, H., Deng, F., Wang, F.: Rough Cluster Algorithm Based on Kernel Function. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 172–179. Springer, Heidelberg (2008)
7. Tripathy, B.K., Ghosh, A., Panda, G.K.: Kernel based K-means clustering using rough set. In: IEEE 2012 International Conference on Computer Communication and Informatics, ICCCI (2012)
8. Tripathy, B.K., Ghosh, A., Panda, G.K.: Adaptive K-Means Clustering to Handle Heterogeneous Data Using Basic Rough Set Theory. In: Meghanathan, N., Chaki, N., Nagamalai, D. (eds.) CCSIT 2012, Part I. LNICST, vol. 84, pp. 193–202. Springer, Heidelberg (2012)
9. Yang, M.S., Tsai, H.S.: A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction. Pattern Recognition Letters 29(12), 1713–1725 (2008)
10. Zhang, D.Q., Chen, S.C.: Kernel Based Fuzzy and Possibilistic C-means Clustering. In: Proc. the International Conference Artificial Neural Network, Turkey, pp. 122–125 (2003)
11. Dubois, D., Prade, H.: Rough Fuzzy sets model. International Journal of General Systems 46(1), 191–208 (1990)
12. Maji, P., Pal, S.K.: RFCM: A Hybrid Clustering Algorithm using rough and fuzzy sets. Fundamenta Informaticae 80(4), 475–496 (2007)
13. Phillips, J.M., Venkatasubramanian, S.: A gentle introduction to the kernel distance, arXiv preprint arXiv:1103.1625 (2011)
14. Mitra, S., Banka, H., Pedrycz, W.: Rough-Fuzzy Collaborative Clustering. IEEE Transactions on System, Man, and Cybernetics, Part B: Cybernetics 36(4), 795–805 (2006)
15. Bezdek, J.C., Pal, N.R.: Some new indexes for cluster validity. IEEE Transaction on System, Man and Cybernetics, Part B: Cybernetics 28, 301–315 (1998)

Palmprint Recognition Using Data Field and PCNN*

Yanxia Wang, Jianmin Zhao, and Guanghua Sun

College of Mathematics, Physics and Information Engineering
Zhejiang Normal University
688 Yingbin Road, Jinhua, Zhejiang Province, 321004 China
{wangyx,zjm,sungh}@zjnu.cn

Abstract. In this paper, an approach is proposed for palmprint recognition, which uses PCNN and data field theory to extract local statistical structure features of a palmprint. In the method, the data field theory is firstly introduced to obtain a relative palmprint data field, which enhances the palm line information. Then the relative data field is input into a PCNN. Next, the local statistical structure features with four values are extracted from each sub-region. At last, all of local statistic-structural feature vectors are weighted and combined into a long feature vector to represent the palmprint. Experiments show that the novel features can effectively characterize different palmprints.

Keywords: Palmprint recognition, data field, PCNN, feature extraction.

1 Introduction

Palmprint recognition uses the features of palmprints to identify a person, that is, principal lines, wrinkles, ridges, minutiae points, singular points and texture etc. [1].In recent years, researchers have proposed many palmprint feature extraction and matching methods, which can be grouped into four categories: structure-based, statistical-based, subspace-based and coding-based [2].

The structure-based methods mainly use the information of the direction and location of principal lines and wrinkles to represent the palmprint [3-4]. These methods are simple and intuitional, but the performance of recognition depends greatly on the edge detectors or broad line detector and several thin and weak palm lines, such as wrinkles, might be too vague for detection by a line detector. The second class approaches include the local statistical-based and global statistical-based methods [5-7]. These methods utilized the transform domain features which can effectively reduce the within class variances to improve the performance. But, in these approached, the size of a block and the strategy of dividing blocks have great

* This work is supported partly by Opening Fund of Top Key Discipline of Computer Software and Theory in Zhejiang Provincial Colleges at Zhejiang Normal University (No. ZSDZZZZXK27), the Zhejiang Province Natural Science Foundations of China (No. LQ13F020026) and the National Natural Science Foundations of China (No. 61272468 and No. 61272130).

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 156–162, 2013.

impact on the recognition results. The third class is subspace-based approaches [8-10]. The subspace coefficients are used to construct feature vectors. Subspace-based methods have successfully applied in palmprint recognition, but subspace is greatly dependent on the training set. Once the training set is subject to change, the whole training procedure will be repeated. The last class methods encode the response of a bank of filters into bitwise code [11-12]. The coding-based method usually has a lower memory requirement and a fast matching speed, and thus has been very successful in palmprint representation and matching, but the methods are sensitive to the translation and rotation of palmprint images.

In order to resolve these problems, we propose a palmprint recognition method of using image data field and PCNN to extract local statistical structure features.

The rest of the paper is organized as follows: A brief description of data field theory is given in Section 2. Section 3 gives a brief overview of PCNN. The proposed method in detail is described in Section 4. Section 5 reports the experimental results and analysis. Finally, the conclusion is given in Section 6.

2 Data Field Theory

In physics, a field is a mode of description of the interaction of material particles. With the concept of the field theory of physics, Li [13] proposed a new field, image data field. In the method, the pixels of an image are seen as the particles with certain mass in the 2D space and the interaction of these pixels makes up an image data field. Let q is a given pixel particle in the image space L, $\forall q \in L$, the potential value of pixel particle q produced on pixel particle p is defined as $\varphi^q(p) = m^q \bullet \exp(-(dist_{pq}/\sigma)^2)$, $dist_{pq} = \max(|x_p - x_q|, |y_p - y_q|)$ .Here, $\sigma > 0$ controls the interaction range of the pixel particle, that is the influence factor; m^q is the intensity of field source; $dist_{pq}$ is the distance between the pixel q and p; x_p, x_q, y_p, y_q are the horizontal and vertical ordinates of the pixel p and q respectively.

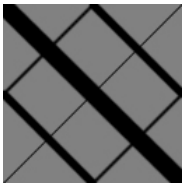
For the data field formed by all of pixel particles in the L, the potential value of any a pixel p can be defined as $\varphi(p) = \sum_{dist_{pq} \le \varepsilon} \varphi^q(p)$.

According to different masses, the absolute image data field and the relative image data field are obtained by Wu [14].If the mass is the difference of grayscales between the center pixel and the pixel influenced by it, then the relative image data field can be obtained:

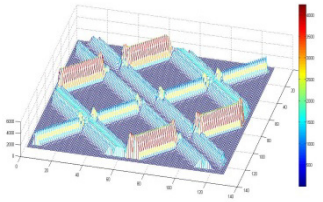
$$\mathrm{m}_r^q = |f(p) - f(q)|, \quad \varphi_r = \sum_{dist_{pq} \le \varepsilon} m_r^q \bullet \exp(-(dist_{pq}/\sigma)^2) \tag{1}$$

Here, m_r^q is the relative mass, respectively; $f(\bullet)$ is the function of the image grayscale. Fig. 1 shows an example of the data field. From the Fig.1, it can be found that the edge information is highlighted by the relative data field. The edge information of

palmlines is what we need, so the relative image data field should be selected to use in our method.



(a)



(b)

Fig. 1. Examples of image data fields. (a) Original image, (b) Relative image data field

3 Pulse-Coupled Neural Network

Pulse-coupled neural network is a called the 3th generation neural network, which is a result of research on artificial neuron model that simulates the behavior of cortical neurons observed in the cortices of animal [15]. The significant advantage of the PCNN model is that it can operate without any pre-train needed. Other than that, it inherits the advantage of artificial neural network (ANN).

The PCNN neuron structure consists of three parts: the receptive field, the modulation field and the pulse generator. The dynamical behavior of neuron can be described as iteration by the following formulas:

$$F_{ij}[n] = S_{ij} + F_{ij}[n-1]e^{-\alpha_F} + V_F \sum_{kl} M_{ijkl} Y_{kl}[n-1] \tag{2}$$

$$L_{ij}[n] = L_{ij}[n-1]e^{-\alpha_L} + V_L \sum_{kl} W_{ijkl} Y_{kl}[n-1] \tag{3}$$

$$U_{ij}[n] = F_{ij}[n](1+\beta L_{ij}[n]), \theta_{ij}[n] = \theta_{ij}[n-1]e^{-\alpha_\theta} + V_\theta Y_{ij}[n-1] \tag{4}$$

$$Y_{ij}[n] = \begin{cases} 1, & U_{ij}[n] > \theta_{ij}[n] \\ 0, & Otherwise \end{cases} \tag{5}$$

Where n is current iteration number. The subscript (i,j) is the position of a neuron in the network and a pixel in the input image. F, L, θ,U and are feeding inputs, linking inputs, dynamic threshold, internal activity and pulse output, respectively. The status of the (i,j) neuron can be indicated by Y_{ij} with binary value. If Y_{ij} equals to 1, then it means the neuron is active. α_F , α_L and α_θ are decay time constants that control how the value of F,L and θ decay over time in the absence of external stimuli. V_F , V_L and V_θ are normalizing constants that scale the inputs from adjacent pixels to their respective fields. M and W are the constant synaptic weight matrices and depended on the

field of the neighborhood neurons. For our implementation, M=W. β is the strength of the linking. S is external input.

4 The Main Idea

In the method, the palmprint image is first mapped from gray space to data field space and a palmprint relative data field is obtained. PCNN is then applied to the entire relative data field. Subsequently, the local statistical structure features with four values are extracted from each sub-region with same size. At last, the all of local statistic-structural feature vectors are combined into a long feature vector to represent the palmprint. The processing flow is outlined in Fig. 2.

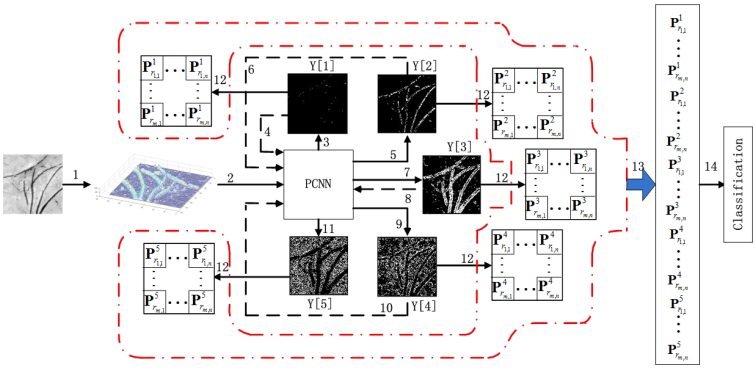


Fig. 2. Processing flow of the proposed method

In order to more effective represent the features of a palmprint, the binary output of each iteration is divided into small sub-regions with 16×16 size. Some statistics and the Tsallis entropy [16] of LBPH are computed from each sub-region and then they are composed of a local statistical structure feature vector, which can be described as:

$$\mathbf{P} = (\mathbf{P}^1_{r_{1,1}}, \mathbf{P}^1_{r_{1,2}}, \cdots, \mathbf{P}^1_{r_{m,n}}, \mathbf{P}^2_{r_{1,1}}, \mathbf{P}^2_{r_{1,2}}, \cdots, \mathbf{P}^2_{r_{m,n}}, \cdots, \mathbf{P}^5_{r_{1,1}}, \mathbf{P}^5_{r_{1,2}}, \cdots, \mathbf{P}^5_{r_{m,n}})$$

$$\mathbf{P}^i_{r_{j,k}} = \left\langle n^i_{r_{j,k}}, \mu^{\ i}_{xr_{j,k}}, \mu^{\ i}_{yr_{j,k}}, T^i_{r_{j,k}} \right\rangle, \quad i = 1,\cdots,5, \quad j = 1,\cdots,m, \quad k = 1,\cdots n \tag{6}$$

$$m = Height_F / Height_r, \quad n = Width_F / Width_r$$

$$n^i_{r_{j,k}} = \sum_{x,y} Y^i_{r_{j,k}}(x,y), \quad Y^i_{r_{j,k}}(x,y) = \{0,1\} \tag{7}$$

$$\mu^{\ i}_{xr_{j,k}} = \sum_{z \in r_{j,k}} x_z \Big/ n^i_{r_{j,k}}, \mu^{\ i}_{yr_{j,k}} = \sum_{z \in r_{j,k}} y_z \Big/ n^i_{r_{j,k}}, \quad Y^i_{r_{j,k}}(z) = 1, \quad z(x,y) \in r_{j,k} \tag{8}$$

$$T^i_{r_{j,k}} = 1 - \sum_{d=0}^{255} (LBPH^i_{r_{j,k}}(d)/255)^q \Big/ (q-1), \quad q > 0 \tag{9}$$

$$LBPH^{i}_{r_{j,k}}(d) = \sum_{x,y \in r_{j,k}} I\left\{LBP^{i}_{r_{j,k}}(x,y) = d\right\}, \quad d = 0,1,\cdots 255 \tag{10}$$

$$LBP^{i}_{r_{j,k}}(x,y) = \sum_{g=0}^{7} Y^{i}_{g}(x,y)2^{g}, \quad Y^{i}_{r_{j,k}}(x,y) \in r^{i}_{j,k}, \quad Y^{i}_{g}(x,y) = \{0,1\} \tag{11}$$

Where $Y^{i}_{r_{j,k}}(x,y)$ is the output of PCNN of the neuron (*x*,*y*) in the $r_{j,k}$ -th sub-region of the i-th iteration. $Height_F$, $Width_F$, $Height_r$ and $Width_r$ are the height and width of relative palmprint field and their sub-regions, respectively. $n^{i}_{r_{j,k}}$ is the number of the neuron of being fired in the $r_{j,k}$ -th sub-region of the *i*-th iteration. $\mu^{i}_{xr_{j,k}}$ and $\mu^{i}_{yr_{j,k}}$ are the mean values of horizontal and vertical coordinates of all of fired neurons, respectively. $T^{i}_{r_{j,k}}$ is the Tsallis entropy of $LBPH^{i}_{r_{j,k}}$. Here, $LBPH^{i}_{r_{j,k}}$ is the local binary pattern histogram extracted from the $r_{j,k}$ -th sub-region of the *i*-th iteration. $LBP^{i}_{r_{j,k}}(x,y)$ is the LBP operation of $Y^{i}_{r_{j,k}}(x,y)$.

5 Experiment Results

5.1 Experiment Data

In this paper, we develop a scanner-based device to capture palm images. The device is used to capture images of both hands from 300 individuals. Ten images are captured from each hand of a person. The center part of each palm is extracted and stored as palmprint images of size 128×128. A total of 6000 images (300×2×10) form a palmprint database.

5.2 Comparisons with other Methods

In the experiment, some methods without any pre-train and our proposed method are tested on our palmprint database. The result of the proposed method is compared with the results of Local Gabor Binary Pattern Histogram Sequence (LGBPHS) method [18], DT-CWT based LBPWH method [10] and PCNN based method without using data field and these results are reported on Table 1.

Compared with the results of different method, it can be found that the recognition rate of the method of only using PCNN is 95.75\%, which is lower than that of other three methods. It because when the enhanced palmprint image is directly input into

Table 1. Recognition rates of different methods

Methods	LGBPHS	DT-CWT based LBPWH	only PCNN	Our method
Results	97.25%	98.35%	95.75%	98.75%

PCNN, there is a white connected area in the output of PCNN. The area produced by the shadow of the palm center may make some useful information be covered and diminish the representation capabilities of their feature vectors. However, palm line information is directly enhanced by introducing relative data field in the proposed method. Therefore, our method gets the better performance than other three methods.

5.3 Speed Test

Speed is another factor in determining the efficiency of an algorithm. The proposed approach and other three methods are implemented using matlab 7.8 on a notebook personal computer with Intel core 2 processor (2.13GHz). Table 2 lists the average execution time of different methods.

Table 2. Execution time of different methods

Methods	LGBPHS	DT-CWT based LBPWH	only PCNN	Our method
Results	0.93s	0.88s	0.81s	0.84s

From the Table 2, we can find that the execution time of all methods is very near. In our method, a relative palmprint data field needs to be constructed as the input of the PCNN. Therefore, our method needs more execution time. Even so, the execution time of our method is 0.84 second, which is fast enough for real-time verification. Thus, considering the various factors, our method applied in palmprint recognition gives the satisfying performance.

6 Conclusion

The paper reports a PCNN based local spatial structure feature extraction method for palmprint recognition. In order to enhance the information of palm lines, the method first introduces data field theory to construct the relative palmprint data field. Moreover, the local statistical structure features are extracted from not the single final output of PCNN but multiple FTMs in different iterations, which make the effective information increase. At last, all of local statistical structure features are composed to a final feature vector. The method is entirely independent on the training set. The proposed method is used to perform accuracy test and speed test. The experimental results demonstrate that the proposed approach can give a better performance. In summary, the PCNN based local statistical structure method is an effective feature extraction approach for palmprint recognition.

References

1. Zhang, D., Wai-Kin, K., You, J., Wong, M.: Online Palmprint Identification. Pattern Analysis and Machine Intelligence 25(9), 1041–1050 (2003)
2. Yue, F., Zuo, W.M., Zhang, D.: Survey of palmprint recognition algorithms. Acta Automatica Sinica 36(3), 353–365 (2010)

3. Leung, F.L.M., Yu, X.: Palmprint matching using line features. In: Proceedings of the 8th International Conference on Advanced Communication Technology, Gangwon-Do, Korea, pp. 1577–1582 (2006)
4. Duta, N., Jain, A.K., Mardia, K.V.: Matching of palmprints. Pattern Recognition Letters 23(4), 477–485 (2007)
5. Kong, W., Zhang, D., Li, W.: Palmprint feature extraction using 2-D Gabor filters. Pattern Recognition 36, 2339–2347 (2003)
6. Jing, X., Zhang, D.: A face and palmprint recognition approach based on discriminant DCT feature extraction. IEEE Transactions on Systems, Man, and Cybernetics, Part B 34(6), 2405–2415 (2004)
7. Wang, Y., Ruan, Q.: Dual-tree Complex Wavelet Transform based Local Binary Pattern Weighted Histogram Method for Palmprint Recognition? Computing and Informatics 28, 299–318 (2009)
8. Lu, G., Zhang, D., Wang, K.: Palmprint recognition using eigenpalms features. Pattern Recognition Letters 24, 1463–1467 (2003)
9. Yang, J., Zhang, D., Yang, J., Niu, B.: Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans. Pattern Anal. Mach. Intell. 29(4), 650–664 (2007)
10. Wang, Y., Ruan, Q.: Kernel Fisher Discriminant Analysis for Palmprint Recognition. In: The Proceedings of the 18th International Conference on Pattern Recognion (ICPR 2006), vol. 4, pp. 457–460 (2006)
11. Kong, A., Zhang, D., Kamel, M.: Palmprint identification using feature-level fusion. Pattern Recog. 39(3), 478–487 (2006)
12. Zhang, D., Zuo, W., Yue, F.: A comparative study of palmprint recognition algorithms. ACM Computing Surveys 44(1), Article 2, 1–37 (2007)
13. Li, D.: Uncertainty in Knowledge Representation. Engineering Science 2(10), 73–79 (2000) (in Chinese)
14. Wu, T., Qin, K.: Image segmentation using cloud model and data field. Pattern Recognition and Artificial Intelligence 25(3), 397–405 (2012)
15. Eckhorn, R., Reitboeck, H.J., et al.: Feature linking via synchronous among distributed assemblies: simulations of results from cat visual cortex. Neural. Comput. 2, 1253–1255 (1990)
16. Zhang, Y., Wu, L.: Pattern recognition via PCNN and Tsallis entropy. Sensors 8, 7518–7529 (2008)
17. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29, 51–59 (1996)
18. Zhang, W., Shan, S., Gao, W., et al.: Local Gabor Binary Pattern histogram Sequence (LGBPHS): A Novel Non-Statistical Model for Face Representation and Recognition. In: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 1, pp. 786–791 (2005)

Highly Sparse Reductions to Kernel Spectral Clustering

Raghvendra Mall, Rocco Langone, and Johan A.K. Suykens

Department of Electrical Engineering, ESAT-SCD, Katholieke Universiteit Leuven,
Kasteelpark Arenberg, 10 B-3001 Leuven, Belgium
{raghvendra.mall,rocco.langone,johan.suykens}@esat.kuleuven.be

Abstract. Kernel spectral clustering is a model-based spectral clustering method formulated in a primal-dual framework. It has a powerful out-of-sample extension property and a model selection procedure based on the balanced line fit criterion. This paper is an improvement of a previous work which sparsified the kernel spectral clustering method using the line structure of the data projections in the eigenspace. However, the previous method works only in the case of well formed and well separated clusters as in other cases the line structure is lost. In this paper, we propose two highly sparse extensions of kernel spectral clustering that can overcome these limitations. For the selection of the reduced set we use the concept of angles between the data projections in the eigenspace. We show the effectiveness and the amount of sparsity obtained by the proposed methods for several synthetic and real world datasets.

1 Introduction

Clustering algorithms are widely used tools in fields like data mining, machine learning, graph compression and many other tasks. The aim of clustering is to divide data into natural groups present in a given dataset. Clusters are defined such that the data present within the group are more similar to each other in comparison to the data between clusters. Spectral clustering methods [1,2,3] are generally better than the traditional k-means techniques. A new spectral clustering algorithm based on weighted kernel principal component analysis (PCA) formulation was proposed in [4]. The method is based on a model built in a primal-dual optimization framework. The model has a powerful out-of-sample extension property which allows to infer cluster affiliation for unseen data.

The data points are projected to the eigenspace and the projections are expressed in terms of non-sparse kernel expansions. In [5] sparsification of this clustering model was done by exploiting the line structure of the projections when the clusters are well formed and well separated. However, *the method fails when the clusters are overlapping and for real world datasets where the projections in the eigenspace do not follow a line structure as mentioned in [8]. In this paper, we propose methods to handle these issues.* We locate the mean of each cluster in the eigenspace. The mean of the cluster in the eigenspace is located on the least squares linear regressor for all the points in that cluster. We use angular

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 163–169, 2013.

distance to locate the projections close to the mean in the eigenspace and then select these projections based on their euclidean distance from the origin in the eigenspace. The main advantage of these sparse reductions is that it results in much simpler and faster predictive models.

2 Kernel Spectral Clustering

We provide a brief description of the kernel spectral clustering methodology. Given a dataset $\mathcal{D} = \{x_i\}_{i=1}^{N_{tr}}, x_i \in \mathbb{R}^d$, the training points are selected by maximizing the quadratic Rènyi criterion as depicted in [8,9,10]. Here x_i represents the i^{th} training point and the training set is represented by X_{tr}. The number of data points in the training set is N_{tr}. Given $\mathcal{D}$ and the number of clusters k, the primal problem of the spectral clustering via weighted kernel PCA is formulated as follows [4]:

$$\begin{aligned} \min_{w^{(l)},e^{(l)},b_l} \quad & \frac{1}{2}\sum_{l=1}^{k-1} w^{(l)\intercal} w^{(l)} - \frac{1}{2N_{tr}}\sum_{l=1}^{k-1}\gamma_l e^{(l)\intercal} D_\Omega^{-1} e^{(l)} \\ \text{such that} \quad & e^{(l)} = \Phi w^{(l)} + b_l 1_{N_{tr}}, l = 1,\ldots,k-1, \end{aligned} \tag{1}$$

where $e^{(l)} = [e_1^{(l)},\ldots,e_{N_{tr}}^{(l)}]^\intercal$ are the projections onto the eigenspace, $l = 1,\ldots,k-1$ indicates the number of score variables required to encode the k clusters, $D_\Omega^{-1} \in \mathbb{R}^{N_{tr}\times N_{tr}}$ is the inverse of the degree matrix associated to the kernel matrix Ω. Φ is the $N_{tr} \times n_h$ feature matrix, $\Phi = [\phi(x_1)^\intercal;\ldots;\phi(x_{N_{tr}})^\intercal]$ and $\gamma_l \in \mathbb{R}^+$ are the regularization constants. We note that $N_{tr} < N$ i.e. the number of points in the training set is less than the total number of points in the dataset. The kernel matrix Ω is obtained by calculating the similarity between each pair of data point in the training set. Each element of Ω, denoted as $\Omega_{ij} = K(x_i, x_j) = \phi(x_i)^\intercal\phi(x_j)$ is obtained for example by using the radial basis function (RBF) kernel. The clustering model is then represented by:

$$e_i^{(l)} = w^{(l)\intercal}\phi(x_i) + b_l, i = 1,\ldots,N_{tr}, \tag{2}$$

where $\phi : \mathbb{R}^d \to \mathbb{R}^{n_h}$ is the mapping to a high-dimensional feature space n_h, b_l are the bias terms, $l = 1,\ldots,k-1$. The projections $e_i^{(l)}$ represent the latent variables of a set of $k-1$ binary cluster indicators given by $\text{sign}(e_i^{(l)})$ which can be combined with the final groups using an encoding/decoding scheme. The decoding consists of comparing the binarized projections w.r.t. codewords in the codebook and assigning cluster membership based on minimal Hamming distance. The dual problem corresponding to this primal formulation is:

$$D_\Omega^{-1} M_D \Omega \alpha^{(l)} = \lambda_l \alpha^{(l)}, \tag{3}$$

where M_D is the centering matrix which is defined as $M_D = \mathrm{I}_{N_{tr}} - (\frac{(1_{N_{tr}} 1_{N_{tr}}^\intercal D_\Omega^{-1})}{1_{N_{tr}}^\intercal D_\Omega^{-1} 1_{N_{tr}}})$. The $\alpha^{(l)}$ are the dual variables and the positive definite kernel function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ plays the role of similarity function. The corresponding predictive model

is $\hat{e}^{(l)}(x) = \sum_{i=1}^{N_{tr}} \alpha_i^{(l)} K(x, x_i) + b_l$ which provides clustering inference for unseen data points x. Thus the model has an out-of-sample extension property. For selection of the hyper-parameters of the model i.e. k and σ for RBF kernel we use the Balanced Angular Fit (BAF) criterion proposed in [6,7].

3 Sparse Reductions to the KSC Model

3.1 Related Work

The projections of the data points in the eigenspace are expressed in terms of non-sparse kernel expansions. The primal vectors $w^{(l)} = \sum_{i=1}^{N_{tr}} \alpha_i^{(l)} \phi(x_i)$ can be approximated by a reduced set. The objective is to approximate $w^{(l)}$ by a new weight vector $\tilde{w}^{(l)} = \sum_{i=1}^{R} \beta_i^{(l)} \phi(\tilde{x}_i)$ minimizing $||w^{(l)} - \tilde{w}^{(l)}||_2^2$ where $\tilde{x}_i$ is the i^{th} point in the reduced set $\mathcal{RS}$ whose cardinality is R. In [5], it was shown that if the reduced set $\mathcal{RS}$ is known then the $\beta^{(l)}$ co-efficients can be obtained by solving the linear system:

$$\Omega^{\psi\psi} \beta^{(l)} = \Omega^{\psi\phi} \alpha^{(l)}, \tag{4}$$

where $\Omega_{mn}^{\psi\psi} = K(\tilde{x}_m, \tilde{x}_n)$, $\Omega_{mi}^{\psi\phi} = K(\tilde{x}_m, x_i)$, $m, n = 1, \ldots, R, i = 1, \ldots, N_{tr}$ and $l = 1, \ldots, k-1$.

This reduced set can be built by selecting points whose projections in the eigenspace occupy certain positions or by using an elastic net penalization. Two methods to attain the same was proposed in [5,8]. The amount of sparsity attained by introducing penalization is not as much as that obtained by selection of points based on their position. The method based on selecting points from certain positions as proposed in [5] works when the clusters are well formed and separated. This is because in that condition we obtain a line structure corresponding to the projections of the data points in the eigenspace.

3.2 Proposed Methods

We propose two methods for selection of points to form the reduced set $\mathcal{RS}$. In case when the clusters are not well formed and overlapping, the projections of the corresponding data points in the eigenspace loses the line structure. We estimate the mean of all the projections for a particular cluster. According to the properties of least squares linear regressor [11], the linear regressor fit for all the projections belonging to that cluster in the eigenspace passes through the mean of the projections. However, it might so happen that in the input space there is no actual data point corresponding to that mean. So, in order to select points existing in the input space, we use the concept of angular similarity.

We select all the projections from a cluster whose cosine distance w.r.t. mean for that cluster (e_{μ_i}) is less than threshold t i.e. $1 - \cos(e_j, e_{\mu_i}) < t$. Here $i = 1, \ldots, k-1$ and $j = 1, \ldots, N_{C_i}$, e_j is a projection and N_{C_i} is the number of points in the i^{th} cluster. We initially set $t = 0$ and increase it using an iterative procedure. During each iteration, we increase the value of the t by 0.1 until we have non-empty selection set corresponding to that cluster. We observe in our experiments that one iteration is enough for most of the datasets.

First Proposed Method - Once the selection set is obtained, we calculate the euclidean distance of these projections from the origin. For the first method of selection, we select the projection which is the farthest and at median distance from the origin. The projection which is the farthest from the origin generally corresponds to a point which is close to the center of the cluster in the input space. This is because the projection value say $e_c^{(l)} = \sum_{i=1}^{N_{tr}} \alpha_i^{(l)} K(x_i, x_c) + b_l$ is dependent on the $\alpha_i^{(l)}$ and the kernel evaluation of $K(x_i, x_c)$. According to the nearly piecewise structure of the eigenvectors, all the points belonging to that cluster have nearly similar value of $\alpha_i^{(l)}$. The cluster center in the input space has maximum similarity (≈ 1) to points in that cluster and minimum similarity (≈ 0) to the points in other clusters. Thus, since the effect of the $\alpha_i^{(l)}$ is nearly constant for all the points belonging to a cluster, it can be concluded that the point corresponding to the cluster center in the input space is the projection whose euclidean distance is farthest from the origin in the eigenspace . The median point is selected to provide more stability to the reduced set $\mathcal{RS}$. Thus, if there are k clusters in a dataset, the number of points required to build the reduced set is $2k$ and the sparsity is given as:

$$Sparsity = 1 - \frac{2k}{N_{tr}} \tag{5}$$

Second Proposed Method - For the second method, we select 10% of the projections from the selection set obtained as a result of cosine distance for each cluster. We keep the vector containing the angular distance between these projections and the corresponding mean in a sorted order. Let the size of this vector be S_i. Then the size of the reduced set for each cluster is $r_i = \lceil \frac{S_i}{10} \rceil$, $i = 1, \ldots, k$. We divide this vector into $\lfloor \frac{S_i}{r_i} \rfloor$ bins for each cluster and select a projection from each bin. The minimum value that sparsity can take for this method is $1 - \frac{N_{tr}}{10 \times N_{tr}}$. This is when all the projections corresponding to each cluster maintain a line structure. This is because the cosine similarity of all the projections w.r.t. mean is maximum. However, in other cases when the projections deviate from the line structure then the amount of sparsity introduced is:

$$Sparsity = 1 - \sum_{i=1}^{k} \frac{1}{N_{tr}} \frac{S_i}{r_i}. \tag{6}$$

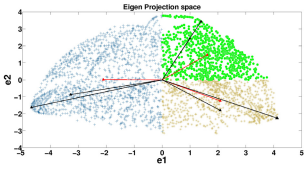
4 Experimental Results

Table 1 provides information about internal quality metrics namely silhouette criterion (*sil*), davies-bouldin index (*db*) and *Sparsity* (*S*) in percentage for the 2 proposed methods. The benchmark datasets are obtained from `http://cs.joensuu.fi/sipu/datasets/`. We compare with method proposed in [5] and the $L_2 + L_1$ penalization based method in [8]. Higher values of *sil* are better and lower values of *db* represents better clustering quality. We highlight the best results in Table 1 for the 9 real world datasets.

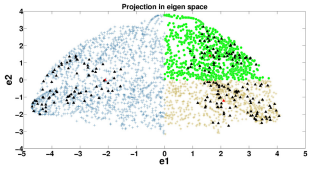
Table 1. Experimental Results on various benchmark datasets

Dataset	1^{st} Highly Sparse Method			2^{nd} Highly Sparse Method			Method in [5]			$L_2 + L_1$ Penalization		
	sil	*db*	*S* (%)	*sil*	*db*	*S* (%)	*sil*	*db*	*S* (%)	*sil*	*db*	*S* (%)
Breast	**0.665**	**0.928**	**98.0**	0.645	0.955	90	0.612	0.975	97.00	0.639	0.933	96.65
Bridge	0.2787	2.05	**99.5**	**0.28**	2.009	91.7	0.265	2.15	99.26	0.275	**1.72**	98.57
Glass	**0.433**	1.85	78	0.35	**1.71**	90	0.32	1.99	69.15	0.41	1.81	**96.88**
Iris	**0.3975**	**0.87**	86.7	0.343	1.12	**91**	0.31	1.25	80.0	0.309	1.306	86.77
MLF	0.74	**1.07**	**99.7**	0.701	1.127	96	0.70	1.15	99.5	**0.795**	1.158	99.55
MLJ	0.823	**0.448**	**99.5**	0.82	0.453	94.8	0.801	0.67	99.2	**0.881**	0.67	95.51
Thyroid	0.499	1.198	**93.75**	**0.512**	**1.183**	90	0.492	1.73	91.0	0.51	1.24	93.75
Wdbc	**0.565**	**1.28**	**97.6**	0.5535	1.28	90	0.54	1.30	95.12	0.56	1.303	**97.6**
Wine	0.282	1.86	88.7	**0.3**	**1.86**	**92.5**	0.273	1.92	82.5	0.29	1.91	86.8

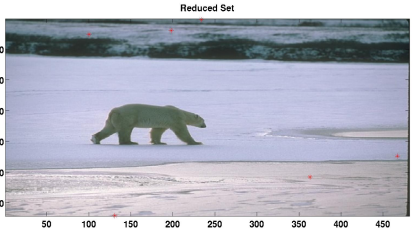
(a) Polar Bear Image



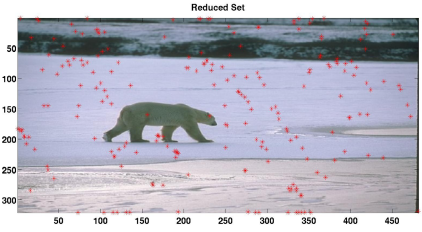
(b) Eigenspace for 1^{st} Highly Sparse Method.



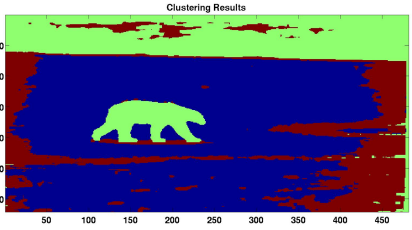
(c) Eigenspace for 2^{nd} Highly Sparse Method.



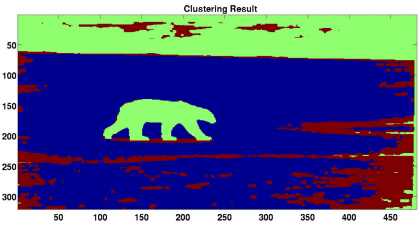
(d) Reduced Set for 1^{st} Highly Sparse Method.



(e) Reduced Set for 2^{nd} Highly Sparse Method.



(f) Clustering Results for 1^{st} Highly Sparse Method.



(g) Clustering Results for 2^{nd} Highly Sparse Method.

Fig. 1. Steps involved for the 2 proposed highly sparse KSC models for an image dataset

An image segmentation experiment using the χ^2 kernel is shown in Figure 1. The total number of pixels is $154,401$ (321×481). The training set consists of $N_{tr} = 7,500$ pixels and the validation set consists of $10,000$ pixels. Both these set are selected by maximizing the quadratic Rènyi entropy. After validation we

obtain $k = 3$ for kernel parameter $\sigma_\chi = 2.807$. The 1^{st} clustering model uses just 6 pixels out of $7,500$ pixels while the second method uses 225 pixels out of $7,500$ training pixels. Since original cluster memberships for this image is not known beforehand, we use 2 internal quality metrics - the silhouette criterion (sil) and the davies-bouldin index (db) as described in [12]. For the 1^{st} highly sparse KSC model the sil value is 0.39 and the db is 1.35 and for the 2^{nd} proposed method these values are 0.32 and 1.15 respectively.

Figure 1a represents the image to be segmented. Figures 1b and 1c showcase the eigenspace. In Figure 1b the red lines represent the cluster means and the black lines represent the farthest and the median projection for that cluster. Similarly, in Figure 1c the red triangles represent the cluster means and the black triangles correspond to 10% of the projections. Figures 1d and 1e highlight the pixels selected from the image as the reduced set $\mathcal{RS}$. These pixels are marked by red-colored '*' reference. Figures 1f and 1g depict the clustering results for the highly sparse KSC models.

5 Conclusion

We proposed 2 highly sparse reductions to KSC model based on a reduced set method. This was achieved by selection of those projections from the eigenspace which occupy certain positions w.r.t. mean projection for each cluster. The clustering model only depended on this reduced set $\mathcal{RS}$ obtained from the training points. This made the clustering model simpler and the predictive model faster as less number of computations were required for out-of-sample extensions. The simulations showed the applicability of the proposed sparse method on various overlapping and real world datasets.

Acknowledgments. This work was supported by Research Council KUL, ERC AdG A-DATADRIVE-B, GOA/10/09MaNet, CoE EF/05/006, FWO G.0588.09, G.0377.12, SBO POM, IUAP P6/04 DYSCO, COST intelliCIS.

References

1. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Dietterich, T.G., Becker, S., Ghahramani, Z. (eds.) Proceedings of the Advances in Neural Information Processing Systems, pp. 849–856. MIT Press, Cambridge (2002)
2. Luxburg, U.: A tutorial on Spectral clustering. Statistics and Computing 17(4), 395–416
3. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Intelligence 22(8), 888–905 (2000)
4. Alzate, C., Suykens, J.A.K.: Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(2), 335–347 (2010)
5. Alzate, C., Suykens, J.A.K.: Highly Sparse Kernel Spectral Clustering with Predictive Out-of-sample extensions. In: ESANN, pp. 235–240 (2010)

6. Mall, R., Langone, R., Suykens, J.A.K.: Kernel Spectral Clustering for Big Data Networks. Entropy 15(5), 1567–1586 (2013)
7. Langone, R., Mall, R., Suykens, J.A.K.: Soft Kernel Spectral Clustering. IJCNN (2013)
8. Alzate, C., Suykens, J.A.K.: Sparse kernel spectral clustering models for large-scale data analysis. Neurocomputing 74(9), 1382–1390 (2011)
9. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific, Singapore (2002)
10. Girolami, M.: Orthogonal series density estimation and the kernel eigenvalue problem. Neural Computation 14(3), 1000–1017 (2002)
11. Kenney, J.F., Keeping, E.S.: Linear Regression and Correlation. Mathematics of Statistics 3(1), ch. 15, 252–285
12. Rabbany, R., Takaffoli, M., Fagnan, J., Zaiane, O.R., Campello, R.J.G.B.: Relative Validity Criteria for Community Mining Algorithms. In: International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 258–265 (2012)

A Rough Clustering Algorithm for Mining Outliers in Categorical Data

N.N.R. Ranga Suri[1], Musti Narasimha Murty[2], and Gopalasamy Athithan[1,3]

[1] Centre for AI and Robotics (CAIR), Bangalore, India
{rangasuri,athithan.g}@gmail.com
[2] Dept of CSA, Indian Institute of Science (IISc), Bangalore, India
mnm@csa.iisc.ernet.in
[3] Presently Working at Scientific Analysis Group (SAG), Delhi, India

Abstract. Outlier detection is an important data mining task with applications in various domains. Mining of outliers in data has to deal with uncertainty regarding the membership of such outlier objects to one of the normal groups (classes) of objects. In this context, a soft computing approach based on rough sets happens to be a better choice to handle such mining tasks. Motivated by this requirement, a novel rough clustering algorithm is proposed here by modifying the basic k-modes algorithm to incorporate the lower and upper approximation properties of rough sets. The proposed algorithm includes the necessary computational steps required for determining the object assignment to various clusters and the modified centroid (mode) computation on categorical data. An experimental evaluation of the proposed rough k-modes algorithm is also presented here to demonstrate its performance in detecting outliers using various benchmark categorical data sets.

Keywords: Data mining, Soft computing, Rough sets, Outlier detection, Data Clustering.

1 Introduction

Rough set theory [7] was basically developed to deal with vagueness in the data. While fuzzy sets deal with such data using a partial membership function, rough sets express the same by the boundary region of a set. A rough set (C) is a set of objects which cannot be with certainty classified as members of the set or its complement using the available knowledge. Thus, associated with every rough set, there is a pair of precise sets known as *lower approximation* ($\underline{C}$) and *upper approximation* ($\overline{C}$) of the rough set. The basic idea is to separate discernible objects from indiscernible ones and to assign them to lower and upper approximations of the set respectively.

As brought out in [1,9], outlier detection is a non-trivial data mining task involving various research issues such as the method of detection, nature of the data, etc. The uncertainty prevailing in crisp labeling of a data object as a normal one or an outlier needs to be handled by employing a soft computing tool

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 170–175, 2013.

such as rough sets. If the underlying data happens to be categorical in nature, as is the case with many real life data mining applications, it leads to more serious concerns due to the non-availability of an effective similarity/dissimilarity measure defined on such data [10].

Motivated by the above discussion, a novel rough clustering algorithm suitable for mining outliers in categorical data is proposed here by modifying the basic k-modes algorithm [4] due to its efficiency as a member of the k-means family of algorithms. The novel algorithm incorporates the lower and upper approximation properties of rough sets similar to the rough k-means algorithm [5].

Section 2 provides an overview of rough clustering including the rough k-means algorithm. The novel rough k-modes algorithm is presented in the subsequent section along with the necessary computational steps. An experimental evaluation of the proposed method for outlier detection is furnished in Section 4. Section 5 concludes this paper with some discussion and ideas for future work.

2 Related Work

In rough clustering, the upper and lower approximations of a cluster are required to satisfy the following basic properties of rough sets [8]:

- A data object can be a member of one lower approximation at most.
- A data object that is a member of the lower approximation of a cluster is also member of the upper approximation of the same cluster.
- A data object that does not belong to any lower approximation is a member of at least two upper approximations.

In contrast to the original rough set theory developed based on the classic set theory, rough clustering is inspired by intervals. The rough k-means algorithm proposed in [5] is an effort in this direction by suitably modifying two important aspects of k-means clustering: (a) calculation of the centroids and (b) object assignment to the lower/upper approximations of a cluster. An initial version of this algorithm was improved upon by proposing various refinements as given in [8]. According to the improved version of the rough k-means algorithm [5], the two fundamental steps mentioned above are computed as follows:

- *Step (A)*: Modified procedure for calculating the centroid Z_j of a cluster C_j.
  - If ($\underline{C_j} \neq \emptyset$ and $\overline{C_j} - \underline{C_j} = \emptyset$) then $z_{j,r} = \frac{\sum_{X_i \in C_j} x_{i,r}}{|\overline{C_j}|}$.
  - If ($\underline{C_j} = \emptyset$ and $\overline{C_j} - \underline{C_j} \neq \emptyset$) then $z_{j,r} = \frac{\sum_{X_i \in (\overline{C_j} - \underline{C_j})} x_{i,r}}{|\overline{C_j} - \underline{C_j}|}$.
  - else $z_{j,r} = w_{low} \frac{\sum_{X_i \in \underline{C_j}} x_{i,r}}{|\underline{C_j}|} + w_{up} \frac{\sum_{X_i \in (\overline{C_j} - \underline{C_j})} x_{i,r}}{|\overline{C_j} - \underline{C_j}|}$.

  where $1 \leq r \leq m$ with m being the dimensionality, and $w_{low} + w_{up} = 1$.
- *Step(B)*: Modified procedure for assigning an object X_i to clusters.
  Let $d(Z_j, X_i)$ be the distance between object X_i and the centroid Z_j of C_j.
  1. Determine the nearest centroid Z_j, s.t. $d(Z_j, X_i) = min_{1 \leq j \leq k} d(Z_j, X_i)$.

2. Determine the centroids Z_l's that are also close to X_i.
   Let $T = \{l : d(Z_l, X_i)/d(Z_j, X_i) \leq \epsilon \text{ and } l \neq j\}$. Then,
   - If $T \neq \emptyset$ then $[X_i \in \overline{C_j} \text{ and } X_i \in \overline{C_l}, \forall l \in T]$.
   - else $[X_i \in \overline{C_j} \text{ and } X_i \in \underline{C_j}]$.

Here, $\epsilon \geq 1$ is a roughness parameter taking a user specified value.

3 Proposed Algorithm

Let $D = \{X_1, X_2, X_3, \ldots, X_n\}$ be the input data set consisting of n data objects, described using m categorical attributes. Each data object X_i is represented as $[x_{i,1}, x_{i,2}, \ldots, x_{i,m}]$. Consider a clustering scenario with k clusters $\{C_1, C_2, \ldots, C_k\}$ represented by the centroids $\{Z_1, Z_2, \ldots, Z_k\}$ respectively.

For the purpose of modifying the basic k-modes algorithm [4] to include the rough set properties, it requires certain measures defined on categorical data to formulate the two important clustering steps mentioned in Section 2. More specifically, a suitable categorical dissimilarity measure between an object X_i and a cluster C_j (with its mode Z_j) is required for assigning the data objects to the clusters. Similarly, computation of new cluster centroids (modes) also needs to be modified by incorporating the frequency counts of various categorical attribute values of the objects assigned to that cluster. In this regard, the following categorical dissimilarity measure as defined in [6] is considered here.

$$d(Z_j, X_i) = \sum_{r=1}^{m} \phi(z_{j,r}, x_{i,r}) \tag{1}$$

where

$$\phi(z_{j,r}, x_{i,r}) = \begin{cases} 1, & \text{if } z_{j,r} \neq x_{i,r}, \\ 1 - \frac{|C_{j,r}|}{|C_j|}, & \text{otherwise.} \end{cases}$$

where $|C_{j,r}|$ is the number of objects with category value $x_{i,r}$ for the r^{th} attribute in the j^{th} cluster.

Let $freq(x_{i,r})$ denote the number of objects in D with the value $x_{i,r}$ for the r^{th} attribute. Then, the density of a data object X_i can be computed as

$$density(X_i) = \frac{1}{mn} \sum_{r=1}^{m} freq(x_{i,r}) \tag{2}$$

Similarly, let $freq_j^{low}(x_{i,r})$ and $freq_j^{up}(x_{i,r})$ denote the number of objects in the lower and upper approximations of j^{th} cluster respectively with the value $x_{i,r}$ for the r^{th} attribute.

Utilizing these categorical measures, various computational steps of the proposed rough k-modes method are formulated as given in Algorithm 1, in accordance with rough set properties listed in Section 2. To ensure faster convergence, the cluster initialization method described in [3] has been considered.

After generating the rough clusters, any clustering-based outlier detection method such as the one in [10], can be employed for producing the outliers.

Algorithm 1. Proposed rough k-modes algorithm

Input: A categorical data set D consisting of n objects of m dimensionality.
Output: k rough clusters (lower and upper approximations) with their modes.

1: Compute $density(X_i)$ of each data object $X_i \in D$ using Equation 2.
2: Determine the initial set of k cluster representatives as per the method in [3].
3: Assign data objects to lower and/or upper approximation of clusters.

- Compute the cluster-to-object distance $d(Z_l, X_i)$ as per Equation 1.
- Determine object assignment according to Step(B) in Section 2.

4: Count cluster-wise attribute value frequencies $freq_j^{low}(x_{i,r})$ and $freq_j^{up}(x_{i,r})$.
5: Compute the new mode $Z_j^* \leftarrow X_i$ of each cluster C_j, s.t. $max_{X_i \in C_j} density_j(X_i)$, where $density_j(X_i)$ is the density of data object X_i w.r.t. cluster C_j given by

- if ($\underline{C_j} \neq \emptyset$ *and* $(\overline{C_j} - \underline{C_j}) = \emptyset$) then $density_j(X_i) = \frac{1}{m} \sum_{r=1}^{m} (\frac{freq_j^{low}(x_{i,r})}{|\underline{C_j}|})$.
- if ($\underline{C_j} = \emptyset$ *and* $\overline{C_j} \neq \emptyset$) then $density_j(X_i) = \frac{1}{m} \sum_{r=1}^{m} (\frac{freq_j^{up}(x_{i,r})}{|\overline{C_j}|})$.
- else $density_j(X_i) = \frac{1}{m} \sum_{r=1}^{m} (w_{low} \frac{freq_j^{low}(x_{i,r})}{|\underline{C_j}|} + w_{up} \frac{freq_j^{up}(x_{i,r}) - freq_j^{low}(x_{i,r})}{|\overline{C_j} - \underline{C_j}|})$. where $w_{low} + w_{up} = 1$.

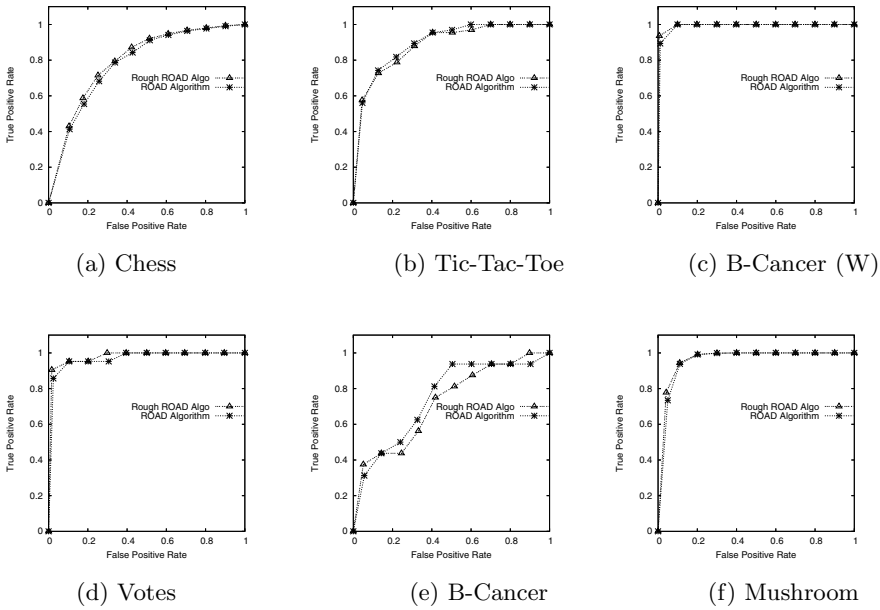
6: Assign data objects to lower and/or upper approximation of clusters as in Step 3.
7: Check for convergence of the algorithm. If not, repeat Steps 4-6 till convergence.

4 Experimental Evaluation

For the purpose of evaluating the effectiveness of the proposed rough k-modes algorithm for outlier detection, a recent related work named as Ranking-based Outlier Analysis and Detection (ROAD) framework [10] has been considered. This framework basically employs two independent ranking schemes to produce a likely set of outliers from the input data. Instead of using the basic k-modes algorithm, the current evaluation employs the novel rough k-modes algorithm with the ROAD framework, hereafter referred to as 'Rough ROAD', for determining the clustering-based ranking of the data objects. The rest of the details regarding this framework remain the same as in [10].

Table 1. Performance comparison on various benchmark data sets

Name of Data Set	Dim	Outlier Class Label	# Normal Objects	# Outlier Objects	# Outliers Detected	
					Rough ROAD Algorithm	ROAD Algo
Chess (End-Game)	36	nowin	1669	305	**131**	126
Tic-Tac-Toe	9	negative	626	66	**38**	37
Breast Cancer (W)	10	4(malignant)	444	47	**44**	42
Congressional Votes	16	republican	124	21	**19**	18
Breast Cancer	9	recurrence-events	196	16	**6**	5
Mushroom	22	poisonous	4208	783	**609**	575



(a) Chess (b) Tic-Tac-Toe (c) B-Cancer (W)

(d) Votes (e) B-Cancer (f) Mushroom

Fig. 1. Comparative view of the performance on various benchmark data sets

This experimental evaluation has been carried out employing six frequently used categorical data sets taken from the UCI ML Repository [2]. A data preparation procedure similar to the one followed in [10] has been applied resulting in the processed data sets as per the details furnished in Table 1.

The evaluation procedure as explained in the ROAD framework has been considered to evaluate the performance of the proposed algorithm. Accordingly, the experimental results obtained using the ROAD methodology with and without the rough k-modes algorithm have been furnished in Table 1. The corresponding ROC curves generated on the benchmark data sets are shown in Fig. 1 for a quick comparison. This evaluation indicates that the proposed rough k-modes clustering algorithm has indeed produced notably better results on all the data sets, highlighting its utility for the outlier detection task.

One of the important parameters of the proposed rough clustering algorithm is ϵ (introduced in Step(B) in Section 2). Varying the value of this parameter has its impact on the rough characteristics of the algorithm, there by on the overall outlier detection performance. To illustrate this impact, further experimentation has been carried out on Mushroom data set [2]. Fig. 2 shows the number of outliers detected on this data set with varying value of this parameter. The other roughness parameters of the algorithm have been set to fixed values as suggested in [8] ($w_{low} = 0.7$, $w_{up} = 0.3$) through out this experimentation and the parameter specific to the ROAD framework as $\alpha = 2$.

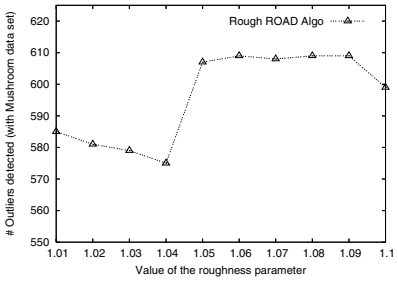


Fig. 2. Outlier detection performance with varying roughness parameter value

5 Conclusion and Future Work

Motivated by the frequent use of rough sets based soft computing approaches in various data mining applications, a novel rough k-modes algorithm has been proposed here for outlier detection. The performance of the proposed algorithm has been experimentally evaluated on various benchmark categorical data sets employing the ROAD framework and also compared it with that of the framework using the basic k-modes algorithm. However, further enhancement to detection accuracy may be possible by experimenting with other roughness parameters.

References

1. Albanese, A., Pal, S.K., Petrosino, A.: Rough sets, kernel set and spatio-temporal outlier detection. IEEE Trans. on Knowledge and Data Engineering (2012) (online)
2. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007), `http://archive.ics.uci.edu/ml`
3. Cao, F., Liang, J., Bai, L.: A new initialization method for categorical data clustering. Expert Systems with Applications 36, 10223–10228 (2009)
4. Huang, Z.: A fast clustering algorithm to cluster very large categorical data sets in data mining. In: SIGMOD DMKD Workshop, pp. 1–8 (1997)
5. Lingras, P., Peters, G.: Applying rough set concepts to clustering. In: Rough Sets: Selected Methods and Applications in Management and Engineering, pp. 23–38. Springer, London (2012)
6. Ng, M.K., Li, M.J., Huang, J.Z., He, Z.: On the impact of dissimilarity measure in k-modes clustering algorithm. IEEE PAMI 29(3), 503–507 (2007)
7. Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)
8. Peters, G.: Some refinements of rough k-means clustering. Pattern Recognition 39, 1481–1491 (2006)
9. Suri, N.N.R.R., Murty, M.N., Athithan, G.: Data mining techniques for outlier detection. In: Visual Analytics and Interactive Technologies: Data, Text and Web Mining Applications, ch. 2, pp. 22–38. IGI Global, New York (2011)
10. Suri, N.N.R.R., Murty, M.N., Athithan, G.: An algorithm for mining outliers in categorical data through ranking. In: IEEE HIS, Pune, India, pp. 247–252 (2012)

On the Use of Monogenic Scale Space for Efficient Face Representation and Recognition

M. Sharmila Kumari[1] and B.H. Shekar[2]

[1] Department of Computer Science and Engineering, P. A. College of Engineering, Mangalore, Karnataka, India
{sharmilabp,bhshekar}@gmail.com
[2] Department of Computer Science, Mangalore University, Karnataka, India

Abstract. In this paper, we present a novel monogenic scale space based Principal Component Analysis (PCA) method by integrating the Reisz transform of face images at different scales and the PCA method for face representation and recognition. The Reisz transform captures desirable facial features characterized by local phase information and local energy at different scales in order to cope with the variations due to illumination and facial pose changes. The PCA method is then employed to reduce the dimension of the feature vectors and hence for efficient face representation and recognition. The feasibility of the proposed monogenic scale space based method integrated with PCA has been successfully tested on many standard face databases such as AT&T and YALE face databases. The recognition accuracy of the proposed approach is compared with the other well known face recognition approaches namely the PCA method, the kernel PCA method and the Gabor wavelet-based RCM method and it is found that the proposed approach exhibit better recognition accuracy when compared to these well known methods.

Keywords: log-Gabor Transform, Riesz transform, Monogenic scale space, Principal component analysis, Face recognition.

1 Introduction

In these days, the biometrics is the well addressed research area in the domain of computer vision and we have seen plethora of algorithms on face based biometrics because of its wide acceptability in several applications ranging from access control, identity authentication, and visual surveillance to human-computer interaction/communication. Devising an efficient face recognition algorithm is quite challenging because of many inherent problems such as inter-class similarity and intra-class variability, occlusion, variations in illuminations, pose changes etc. In order to address these problems, efforts are made by many researchers that result in different classes of algorithms. The local descriptor based algorithms gain much importance because of their robustness to noise and occlusion-the common problems encountered in real recognition environment. The transform based algorithms are most useful in face recognition as they are capable of withstanding illumination problems.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 176–183, 2013.

The local descriptors [8], [14], [22] are commonly employed in a number of real-world applications such as object recognition [5], [14] and image retrieval [16] as they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint. Mikolaczyk and Schmid [15] presented a comparative study of several local descriptors including steerable filters [7], differential invariants [10], moment invariants [22], Scale Invariant Feature Transform [13], and cross-correlation of different types of interest points [8], [16]. Their experiments showed that the ranking of accuracy for the different algorithms was relatively insensitive to the method employed to find interest points in the image but was dependent on the representation used to model the image patch around the interest point. In the SIFT algorithm, each keypoint is represented by its neighborhood, described as a set of orientation histograms computed from the gradient image. The SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change (in a narrow range). The most common implementation uses 16 histograms of 8 bins (8 orientations), which gives a 128 dimensional descriptor. Ke and Sukthankar [9] proposed PCA-SIFT descriptor which is also based on the gradient image, the main difference with SIFT being the further compression using PCA. Recently, the Speeded Up Robust Features (SURF) descriptor [1] has appeared as an alternative to SIFT. Its main advantage is its fastest computation, while keeping a high descriptive power. It is partially inspired by SIFT, but instead of using the gradient image, it computes first order Haar wavelet responses. The Local Energy based Shape Histogram (LESH) has been specifically designed for face recognition applications. Its goal is to encode the underlying shape present in the image. Basically, the descriptor is a concatenation of histograms obtained by accumulating local energy at several filter orientations.

On the other hand, we have seen appearance based approaches which are capable of withstanding noise and illumination and simple in terms of implementation. In these approaches, data transformation is a fundamental step and the goal is to obtain highly discriminative lower-dimensional data from high-dimensional data. Principal component analysis (PCA) and linear discriminate analysis (LDA) are the widely used techniques in the face recognition domain, which encode high-dimensional face images as lower diimensional eigenfaces [20] and fisherfaces [2] respectively. PCA is a linear method that ensures that the data transformed are uncorrelated and preserve maximally the second order statistics of the original data, and hence is insensitive to the dependencies of multiple features in the patterns. To overcome this problem, kernel PCA [19] is proposed as a non-linear extension of PCA that computes the principal components in a high-dimensional feature space. On the similar line, kernel FLD is proposed as a non-linear extension to linear discriminant analysis.

We have also seen the frequency domain based approaches where high frequency components are used as facial features because of their robustness to illumination changes. The local phase information is used in some techniques as it is proved to be sufficient to completely reconstruct a signal within a scale factor [17]. As the Gabor wavelet representation captures salient visual properties

such as spatial localization, orientation selectivity, Lades et al [11] applied Gabor wavelets for face recognition using dynamic link architecture framework. Wiskott et al. [23] extended this basic frame work to devise Gabor wavelet based elastic bunch graph matching method to label and recognize faces. Liu [12] proposed Gabor-based kernel PCA method by integrating the Gabor wavelet representation of face images and the kernel PCA method for face recognition. Porikli and Tuzel [18] and Tuzel et al.[12] proposed a new descriptor framework called region covariance matrices (RCM) for object detection and tracking. The RCMs can be categorized as a matrix-form feature. However, direct application of RCM to face recognition has not produced satisfactory results and hence Yanwei et al., [24] introduced Gabor-based region covariance matrices as face descriptors. The results of Yanwei et al. [24] are highly encouraging when compared to any other Gabor wavelet based techniques. Although Gabor wavelet based face recognition models possess very high recognition accuracy irrespective of noise, occlusion and illumination problems, massive computing and space requirements are the major bottleneck in these approaches. In this context, we have proposed a new frame work for face recognition based on monogenic scale space approach. The monogenic signal requires fewer convolutions when compared to Gabor wavelet based model and hence require much lesser computing time. The principal component analysis is employed on monogenic scale space based images to obtain compact representation of face images.

The remaining part of the paper is organized as follows. In section 2, we discussed monogenic scale space and its application in the context of face recognition is analyzed. The PCA based face representation in monogenic scale space is given in section 3. The experimental results and comparative study is given in section 4. Conclusion is reached in section 5.

2 Monogenic Scale Space: A Review

The monogenic signal [3], [4] is based on the Riesz transform which is used instead of the Hilbert transform. The monogenic signal analysis [3], [4] is a framework to interpret images in terms of the local phase, local orientation and local energy. The monogenic signal is an effective tool to analyze 2-D signals in a rotation invariant manner. The signal is built upon the first order Riesz transform. The spatial representation of the Riesz kernel in 2D space is:

$$(R_x(\mathbf{x}), R_y(\mathbf{x})) = \left(\frac{x}{2\pi |\mathbf{x}|^3}, \frac{y}{2\pi |\mathbf{x}|^3} \right), \mathbf{x} = (x, y) \in \mathbb{R}^2 \tag{1}$$

and its transfer function in the Fourier domain is:

$$(F_u(\mathbf{u}), F_v(\mathbf{u})) = \left(-i\frac{u}{|\mathbf{u}|}, -i\frac{v}{|\mathbf{u}|} \right), \mathbf{u} = (u, v) \in \mathbb{R}^2 \tag{2}$$

For any image, say $I(x)$, the monogenic signal is defined as the combination of I and its Riesz transform:

$$i.e.\ I_m(\mathbf{x}) = (I(\mathbf{x}), R_x\{I\}(\mathbf{x}), R_y\{I\}(\mathbf{x}))$$

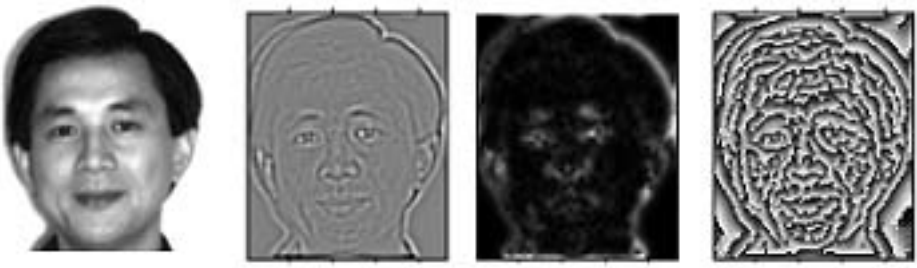


Fig. 1. Original image; log-Gabor image; Energy image and Orientation image

$$= (I, R_x * I, R_y * I) \tag{3}$$

where $*$ stands for the convolution operation. And hence the local orientation is calculated as [3, 4]:

$$\theta = \arctan \frac{R_y \{\mathbf{I}\}}{R_x \{\mathbf{I}\}}, \theta \in [0, \pi) \tag{4}$$

The local phase is defined as:

$$\phi = \arctan 2 \left(\sqrt{R_x^2 \{\mathbf{I}\} + R_y^2 \{\mathbf{I}\}}, \mathbf{I} \right), \phi \in [0, \pi) \tag{5}$$

The local energy is defined as:

$$E = \left(\sqrt{R^2 \{\mathbf{I}\} + R_x^2 \{\mathbf{I}\} + R_y^2 \{\mathbf{I}\}}, \mathbf{I} \right) \tag{6}$$

where $R^2 \{\mathbf{I}\} = I * F^{-1}(G(w))$.Here $G(w)$ is the log-Gabor filter in the Fourier domain. Since log-Gabor filters are band-pass filters, usually multi-scale monogenic representation is required to fully describe a signal. In Fig 1, we have given the convolved images in monogenic scale space filter showing the log-Gabor transformed image with its energy and orientation images. One can see that the local structure is well captured in monogenic components.

3 Face Representation Using PCA in Monogenic Scale Space

Let there be N number of training images. Let A_i, i=1...N, be an image of size m x n. Let each image A_i, i=1...N, has p number of local phase and local energy captured images. These 2p number of images are concatenated to form 2pmxn dimensional feature vector. Hence, we are having $2p(mxn)xN$ sized training matrix called U. Let $\overline{A}$ be the average image of all the images: 2p x N. Let U be made as mean centred. i.e.,

$$\boldsymbol{U} = [U_{1,1} - \overline{A}, U_{1,2} - \overline{A}, ..., U_{1,K_N} - \overline{A}] \tag{7}$$

where $U_{1,1}$ represents the first training image obtained by both local phase and local energy at different scales,, $U_{2,1}$ represents the second training image obtained by both local phase and local energy at different scales and so on. The training matrix $\boldsymbol{U}$ given in Eq. (7) contains *2pmn*-dimensional feature vector which are mean centered. Unlike appearance based models where intensity values are used as feature vectors, here we have considered local energy and local phase information associated with each image as a feature vector. However, a large number of training samples are often needed to get reliable and robust estimation about the characteristics of data distribution which is often called as *curse of dimensionality*. The eigenvectors, e_i and the corresponding eigenvalues λ_i of $\boldsymbol{U}$ are determined by solving the well-known eigen-structure decomposition problem:

$$\lambda_i e_i = \boldsymbol{U} e_i \tag{8}$$

Though all the eigenvectors are needed for accurate recognition, only a small number, q is generally sufficient for capturing the primary characteristics of the feature descriptors. The q eigenvectors, corresponding to the q largest eigenvalues, constitute the monogenic eigenspace. Thus monogenic eigenspace analysis can drastically reduce the dimension *2pmn* to the monogenic eigenspace dimension q while keeping several of the most effective features that summarize the original information.

3.1 Feature Extraction

The optimal projection axes: $\lambda_1, \lambda_2, ..., \lambda_q$ are used for feature extraction. Given an image A, compute the local phase and local energy images, S of the image A and project the local phase and local energy images onto the optimal projection axes λ_s that results in a feature matrix. That is,

$$F = S\lambda_j, \forall j = 1..q \tag{9}$$

The above process of projection of training images onto monogenic eigenspace is to be repeated for all the training images to create a knowledge base.

3.2 Face Recognition

Let I be an image given for recognition. Let $I^{'}$ be the local phase and local energy captured image which is projected onto the q number of optimal projection axes λ_s that results in test image feature matrix say, T computed by $T = I^{'}\lambda$. Given two images, say $\gamma_{i_1} and \gamma_{i_2}$ of any two face(s), represented by feature vectors: $r = \lfloor r_1, r_2, ..., r_q \rfloor$ and $s = \lfloor s_1, s_2, ..., s_q \rfloor$, the *dist(r,s)* is defined as

$$dist(r, s) = \sum_{j=1}^{q} \|r_j - s_j\| \tag{10}$$

where $\|a - b\|_2$ denotes the Euclidean distance between the two vectors a and b. For classifying a given test image, the nearest neighbour is identified among the training samples and the corresponding class label of nearest neighbour training sample is attached to the image under testing.

4 Experimental Results

This section presents the results of the experiments conducted to corroborate the success of the proposed model. We have conducted experimentation on AT&T and YALE face datasets. We have specifically chosen this dataset as these are used by many researchers as a benchmark dataset to verify the validity of their proposed face recognition models. All experiments are performed on a P-IV 2.99GHz Windows machine with 8GB of RAM.

Experimentation on AT&T face dataset: The AT&T face dataset contains images from 40 individuals, each providing 10 different images of size 112x92. In our experiment, we have considered alternate five samples per class for training and the remaining samples for testing. Similarly, we have conducted experiments considering 160 faces as training faces of the AT&T database choosing remaining 6 faces from each person as test faces and the recognition performance has been obtained considering the remaining faces as test faces. The recognition accuracy of the proposed approach with varying dimension of feature vectors is given in Table 1.

Table 1. Recognition accuracy of the proposed approach and other techniques on AT&T Face database

Algorithm	No. of Training samples	Dimension of featuer vector				
		30	35	40	45	50
Eigenface	200	88	87.5	86.5	86.5	86.5
	160	83.33	80.41	80.41	81.25	81.25
Kernel Eigenface	200	88	88	88	88	88
	160	87.5	87.5	87.5	87.5	87.5
Proposed approach	**200**	**94.5**	**95**	**96**	**96**	**96.5**
	160	**95**	**95**	**95**	**95.41**	**95.41**
Gabor Wavelet	200	86.50 – Here the dimension of the feature vector is : 1600				
	160	84.00 – Here the dimension of the feature vector is : 1600				

Table 2. Recognition accuracy of the proposed approach and other techniques on YALE Face database

Algorithm	No. of Training samples	Dimension of featuer vector				
		30	35	40	45	50
Eigenface	75	86.66	86.66	86.66	86.66	86.66
	90	82.66	82.66	84	84	82.66
Kernel Eigenface	75	85.55	85.55	85.55	85.55	85.55
	90	82.66	82.66	82.66	82.66	82.66
Proposed approach	**75**	**91.11**	**92.22**	**92.22**	**92.22**	**92.22**
	90	**89.33**	**89.33**	**90.66**	**90.66**	**90.66**
Gabor Wavelet	75	85.89 – Here the dimension of the feature vector is : 1600				
	90	84.30 – Here the dimension of the feature vector is : 1600				

Experimentation on YALE face dataset:The YALE face dataset contains 165 images of 15 subjects that include variation in both facial expression and lighting. The training set comprised of five and six images randomly chosen for each person with remaining number of face images for each person. The recognition accuracy of the proposed approach with varying dimension of feature vectors is given in Table 2. In Tables 1 and 2, we have also provided the recognition accuracy of the standard eigenface approach, kernel PCA and Gabor wavelet based RCM techniques. It shall be observed from Table 1 and 2 that the proposed approach possess best recognition accuracy when compared to the existing approaches.

5 Conclusion

We have presented a new face recognition algorithm within the frame work of monogenic scale space that captures the local energy and local phase information at different scales. For the compact representation of the phase and energy contained images, principal component analysis is employed and there by crated the knowledge base. Experimental results on the standard benchmark face databases and comparative study with the well known face recognition approaches reveal the superiority of the proposed approach for face recognition problems and its suitability in real environment.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
2. Belhumeur, P.N., Hespanha, J.P., Kreigman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
3. Felsberg, M., Sommer, G.: The monogenic scale-space: A unifying approach to phase-based image processing in scalespace. Journal of Mathematical Imaging and Vision 21(1), 5–26 (2004)
4. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Trans. SP 49, 3136–3144 (2001)
5. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: Proceedings of Computer Vision and Pattern Recognition (June 2003)
6. Forstner, W., Moonen, B.: A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University (1999)
7. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Trans. Pattern Analysis and Machine Intelligence 13(9), 891–906 (1991)
8. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, pp. 147–151 (1988)
9. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, Washington (2004)

10. Koenderink, J., van Doorn, A.: Representation of local geometry in the visual system. Biological Cybernetics 55, 367–375 (1987)
11. Lades, M., Vorbruggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion Invariant Object Recognition in the Dynamic Link Architecture. IEEE Trans. Computers 42, 300–311 (1993)
12. Liu, C.: Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5) (May 2004)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004)
14. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of International Conference on Computer Vision, pp. 1150–1157 (1999)
15. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: Proceedings of Computer Vision and Pattern Recognition (June 2003)
16. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: Proceedings of International Conference on Computer Vision, pp. 525–531 (July 2001)
17. Oppenheim, A.V., Lim, J.S.: The importance of phase in signals. Proceedings of the IEEE 69(5) (1981)
18. Poriklim, F., Tuzel, O.: Fast construction of covariance matrices for arbitrary size image windows. In: Proceedings of IEEE International Conference on Image Processing, pp. 1581–1584 (2006)
19. Schokopf, B., Mika, S., Burges, C.J.C., Knirsch, P., Muller, K.-R., Ratsch, G., Mola, A.J.: Input space versus feature space in kernel based methods. IEEE Transactions on Neural Networks 10(5), 1299–1319 (1999)
20. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
21. Tuzel, O., Porikli, F., Meer, P.: Region covariance: a fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
22. Van Gool, L., Moons, T., Ungureanu, D.: Affine/photometric invariants for planar intensity patterns. In: Buxton, B.F., Cipolla, R. (eds.) ECCV 1996. LNCS, vol. 1064, pp. 642–651. Springer, Heidelberg (1996)
23. Wiskott, L., Fellous, J.M., Kruger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 775–779 (1997)
24. Yanwei, P., Yuan, Y., Li, X.: Gabor-based region covariance matrices for face recognition. IEEE Transactions on Circuits and Systems for Video Technology 18(7) (2008)

Fingerprint Recognition Based on Adaptive Neuro-Fuzzy Inference System

Tripti Rani Borah, Kandarpa Kumar Sarma, and Pran Hari Talukdar

[1] Department of Computer Science, Gauhati University Guwahati, Assam, India
{tbtriptidas29,kandarpaks}@gmail.com
[2] Dept. of Electronics and Communication Technology, Gauhati University Guwahati, Assam, India
[3] Department of Instrumentation and USIC, Gauhati University Guwahati, Assam, India
phtalukdar@gauhati.ac.in

Abstract. Fuzzy logic (FL) is a powerful problem solving methodology receiving wide spread acceptance for a range of applications. FL is also considered for image understanding applications such as edge detection, feature extraction, classification and clustering. It provides a simple and easy way to draw a definite conclusion from ambiguous, imprecise or vague information. Like Artificial Neural Network (ANN) models, some fuzzy inference system (FIS)s have the capability of universal approximation. The adaptive neuro-fuzzy inference system (ANFIS) belongs to the class of systems commonly known as neuro-fuzzy systems (NFs). NFs combines the advantages of ANN with those of fuzzy systems. An ANFIS based identification system is described here which uses fingerprint as an input. Experiments are carried out using a number of samples. Obtained results show that the system is reliable enough for considering it as a part of a verification mechanism.

1 Introduction

To deal with security problems, the use of biometric-based technology (specifically through fingerprint, face, speech etc.) are considered, because such biometric factors have unique characteristics and are related to specific detection and decision states. The related processing technology has reached a high degree of maturity, such as allowing successful application on secure authentication [1]. Fingerprints have been used for over a century and are the most widely used form of biometric identification. The fingerprint of an individual is unique and remains unchanged over a lifetime [2]. Fingerprints are graphical flow-like ridges and valleys present on the surface of human fingers. Typically, there are two prominent types of minutiae (ridge endings and ridge bifurcations) that constitute a fingerprint pattern. A fingerprint can be represented by the minutiae locations, types and attributes like orientation. A good quality fingerprint image typically has about 40 to 100 minutiae, but a dozen of minutiae are considered sufficient to identify a fingerprint pattern. An fingerprint recognition system (FRS) is one of the popular biometric techniques and is an automated method of verifying a match between fingerprint images [3]. Though fingerprint of an individual is unique and doesnot change over a lifetime, for a large number of such samples, usually seen in a biometric authentication

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 184–189, 2013.

system, the classifier must have the ability to capture and retain the minute variations observed in the applied patterns. Fuzzy systems are useful for dealing with uncertainty. Fuzzy logic (FL) is a powerful problem solving methodology receiving wide spread acceptance for a range of applications. FL is also considered for image understanding applications such as edge detection, feature extraction, classification and clustering. It provides a simple and easy way to draw a definite conclusion from ambiguous, imprecise or vague information. Like Artificial Neural Network (ANN) models, some fuzzy inference system (FIS)s have the capability of universal approximation. The adaptive neuro-fuzzy inference system (ANFIS) belongs to the class of systems commonly known as neuro-fuzzy systems (NFs). NFs combines the advantages of ANN with those of fuzzy systems. The ANFIS combines the advantages of neuro and fuzzy based systems and proves to be adaptive and robust in dealing with finite variations. Here, we propose a fingerprint recognition system (FRS) based on ANFIS. Experimental results show that the proposed approach is fast, reliable and robust for a range of conditions. An ANFIS can be configured and trained to handle range of variations observed in the texture of the fingerprint. The specialty of the work is associated with the fact that if the ANFIS is configured properly it can tackle the variations in the fingerprint images. This way the approach provides the insights for developing a system which requires these samples for verification and authentication. Some of the related works are [7] to [8].

The rest of the paper is organized as follows. Section 2 provides the proposed system model and implementation details. All experimental results and related discussions are provided in Section 3. This paper is concluded by summing up the work in Section 4.

2 System Model and Implementation Details

The ANFIS has a processing system which requires sub-blocks of fuzzification (done using membership functions (MF)), fuzzy rules for inference generation and defuzzification. The ANFIS combines the advantages of ANN with those of fuzzy systems. Here, we consider Sugeno fuzzy model [4]. A generic FRS using ANFIS in block diagram form is shown in Figure 1. The first module is image acquisition which is followed by preprocessing. The feature extraction stage enables the capture of vital information about the input samples. These are next fuzzified using certain MFs. Classification stage is the key component of the system and determines the system's performance to a large extent. ANFIS is used as classifier and it produces the correct result by classifying the feature extracted templates and matching these features with known patterns in the feature database. An ANFIS basic building block diagram is shown in Figure 2 with its five layers of different functions which constitute the system. The first layer consists all the adaptive nodes with membership values as outputs which is followed by the next layer consisting of the fixed node which multiplies the incoming signals and gives the product as output. The next layer is the normalization layer. The fourth layer is an adaptive node which is followed by the fifth layer with the overall output of the model.

In this proposed model, a multi stage approach is used. Fingerprint images are captured by a digital scanner. Preprocessing includes resizing images, normalization, segmentation and thinning etc. The next stage is feature extraction. The thinned images are considered for the minutiae feature extraction. The minutiae feature extraction algorithm extracts the minutiae features required for our purpose. Here, Crossing Number

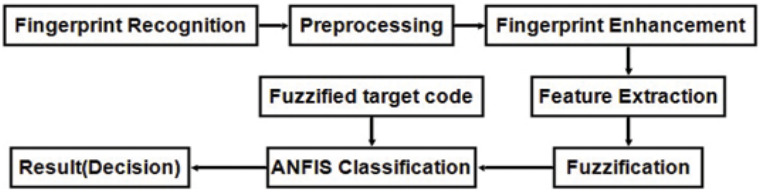


Fig. 1. Process logic of the complete system

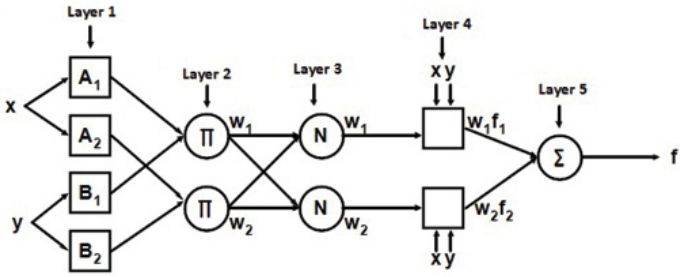


Fig. 2. Building blocks of ANFIS

(CN) method is used for minutiae extraction of fingerprints [5]. The ridge pixel can be divided into bifurcation, ridge ending and non-minutiae point based on it. The CN algorithm is working on pixel representation 1 or 0 but the decision of minutiae point can be selected for each pixel value. CN method extracts the ridge endings and bifurcations from the skeleton image by examining the local neighborhood of each ridge pixel using a 3×3 window. The CN for a ridge pixel P is given by

$$CN = 0.5\sum_{i=1}^{8}|P_i - P_{i+1}|, \quad P_9 = P_1 \tag{1}$$

where P_i is the pixel value in the neighborhood of P. For a pixel P, its eight neighboring pixels are scanned in an anti-clockwise direction.

After the CN for a ridge pixel has been computed, the pixel can then be classified according to the property of its CN value. With this formula, if CN=1 it corresponds to the end point and if CN=3, it corresponds to bifurcation point of minutiae. Other properties of CN are described in Table 1. While applying this algorithm, border area may be ignored, since there is no need to extract minutiae point on border area of the image that will gives more false minutiae points. After a successful extraction of minutiae, they are stored in a template, which may contain the position, direction, type and quality of the minutiae. Here, ANFIS is used as classifier for recognition. It applies a combination of the least-squares (LS) method and the back propagation gradient descent (BPGD) method for training FIS membership function parameters to emulate a given training data set. Fingerprint with CN features of length 101 is used for the training of ANFIS for 10 to 60 epochs. The results obtained are average values of atleast ten

Table 1. Properties of CN

CN	Properties
0	Isolated Point
1	Ending Point
2	Connective Point
3	Bifurcation Point
4	Crossing Point

trails for the epochs considered. We have used the generalized bell-shaped MF for doing the fuzzification. This is because the bell-shaped MF provides the most reliable real world to fuzzy conversion as has been known from experiments. We have considered ten such MFs which gives optimum results. Also, we consider the root mean squared error (RMSE) to be the cost functional for training the ANFIS. We consider about forty clean fingerprint samples. From them, a few set of noise corrupted samples are also formed for testing.

Table 2. ANFIS Specifications

Input Data Size	Fingerprint with CN features of length 101
SNR	0 to 3db
ANFIS type	ANFIS with five layers
ANFIS training method	Least-squares method with the back propagation gradient descent method
Average training epochs	10 to 60
Total number of membership functions	10
Type of membership function	Generalized bell-shaped built-in membership function
RMSE with 3 MFs at 60 epoch	0.01
RMSE with 10 MFs at 50 epoch	0.005

Table 3. Average success rates achieved between a few numbers of training epochs with ANFIS and ANN

Observation No.	Epochs by ANFIS	Epochs by ANN	%Success rate with ANFIS	%Success rate with ANN
1	10	15	88	86
2	20	32	92	90
3	30	47	94.5	92
4	40	61	95.2	93
5	50	76	96.5	94
6	60	91	98	96

3 Experimental Details and Results

The performance of FRS is analyzed in terms of computational speed and reliability. The overall computational time taken by the system is reduced to a greater level. The obtained results are compared with [6] and found that the ANFIS generate comparable RMSE values but the computational speed is significantly better. The ANFIS shows at least 50% increase in computational speed which establishes its usefulness. Further, the RMSE convergence is completed in less number of epochs. Thus the proposed approach is suitable for verification and authorization for the design of FRS. A total of 40 identical fingerprint images have been provided to the system for training, validation and testing of the system. After extensive training, the system is subjected to certain variations with signal to noise ratio (SNR) ranging between 0 to 3 dB to achieve robustness and proper recognition. The ANFIS considered is configured using the specifications shown in Table 2. The average RMSE convergence plot shown by the ANFIS during training while configuring the FRS is shown in Figure 3. The epochs are between ten to sixty and the success rate is around 88 to 98%. The training time required is between 8 to 10 seconds for a set of ten samples each. The results are derived by performing ten trials for each of the sample sets and the average results are quoted.

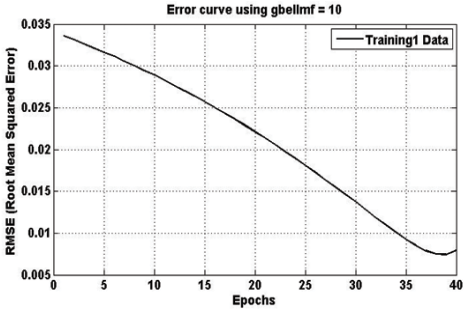


Fig. 3. Average RMSE convergence of the ANFIS upto 40 epochs with 10 MFs

4 Conclusion

Here, we described a FRS where the ANFIS forms a critical decision support system. The experimental results show that the proposed approach is at least 50% more efficient compared to ANN based approaches reported earlier. The strength of the proposed system is its speed, computational efficiency, robustness and high precision which shall make it suitable for certain application. The system proposed here is reliable and efficient enough to be a part of a biometric verification system. The future scope of this work is to enhance the ANFIS architecture to achieve high classification accuracy. The overall performance of the system can be enhanced further by considering more number of samples and variations and by using of better preprocessing techniques with ANFIS based blocks.

References

1. Jain, A., Hong, L., Pankanti, S.: Biometric Identification. Communications of the ACM 43(2), 91–98 (2000)
2. Jain, A., Hong, L., Pankanti, S., Bolle, R.: An identity authentication system using fingerprints. Proceedings of the IEEE 85, 1365–1388 (1997)
3. Thai, R.: Fingerprint Image Enhancement and Minutiae Extraction. B. Eng. thesis, The University of Western Australia (2003)
4. Ross, T.J.: Fuzzy logic with Engineering Applications, 2nd edn. Wiley India, New Delhi (2008)
5. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition, 2nd edn. Springer Professional Computing, Berlin (2009)
6. Borah, T.R., Sarma, K.K., Talukdar, P.H.: Fingerprint Recognition using Artificial Neural Network. International Journal of Electronics Signals and Systems (IJESS) 3(1), 98–101 (2013)
7. Jang, R.J.: ANFIS: Adaptive-Network Based Fuzzy Inference System. IEEE Transaction on Systems, Man and Cybernetics 23(3) (1993)
8. Hsieh, C.T., Hu, C.S.: An Application of Fuzzy Logic and Neural network to Fingerprint Recognition. In: Proceedings of IEEE-Eurasip Nonlinear Signal and Image Processing (NSIP 2005), Japan, vol. 20 (2005)

Dynamic Programming for Bayesian Logistic Regression Learning under Concept Drift

Pavel Turkov[1], Olga Krasotkina[1], and Vadim Mottl[2,⋆]

[1] Tula State University, Tula, Russia
{pavel-turkov,o.v.krasotkina,vmottl}@yandex.ru
[2] Computing Center of the Russian Academy of Sciences, Moscow, Russia

Abstract. A data stream is an ordered sequence of training instances arriving at a rate that does not permit to permanently store them in memory and leads to the necessity of online learning methods when trying to predict some hidden target variable. In addition, concept drift often occurs, what means that the statistical properties of the target variable may change over time. In this paper, we present a framework of solving the online pattern recognition problem in data streams under concept drift. The framework is based on the application of the Bayesian approach to the probabilistic pattern recognition model in terms of logistic regression, hidden Markov model and dynamic programming.

Keywords: online learning, concept drift, logistic regression, hidden Markov model.

1 Introduction

A data stream is an ordered sequence of instances that arrive at a rate that does not permit to permanently store them in memory [1]. Unfortunately the standard approach to pattern recognition learning is caused by the tacit assumption that the entire available information is to be processed at once as a single data chunk. However, in the case of a data stream, such a traditional data mining concept would require infinite storage and running time. The data stream leads to the necessity of online learning methods, which assume the updating of an existing classifier or pool of classifiers. The main goal of online learning is to incorporate the knowledge that is intrinsically present in previously observed chunks of data into the existing system.

Pioneer researches concentrated on the creation of algorithms for implementation of online learning with respect to different types of classifiers. The majority of them considered a single particular classifier and confined themselves to a direct tuning of its parameters [2,3]. Many authors proposed one-pass versions of batch learning algorithms. Later, ensembles of classifiers were additionally applied to the problems which had been originally assumed to be solved by single classifier-based incremental learning algorithms [1], [4,5].

⋆ We would like to acknowledge support from grants of the Russian Foundation for Basic Research 11-07-00409, 11-07-00634 and 12-01-31524.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 190–195, 2013.

In some applications the target concept may change while being analyzed, in the literature such a problem is known as the concept drift problem. There are some methods for pattern recognition under concept drift [6], that use a sliding window to choose a group of new instances to train a model. Such methods include a single classifier like [7,8] or exploit an ensemble of classifiers [9].

However, it is to be noted that the accurate mathematical statement of concept drift does not exist. The known algorithms are somewhat heuristic, and each specific heuristics is determined by the specificity of the particular practical task.

In this paper, we propose a probabilistic Bayesian approach to the online pattern recognition problem based on treating the concept drift as a hidden Markov process and exploiting the general principle of dynamic programming.

2 Background

Let each of real-world objects hiddenly belong to one of two classes $y = \pm 1$ and be accessible to immediate observation only through its real-valued feature vector $\mathbf{x} \in \mathbb{R}^n$. The classical linear approach to the pattern recognition problem is underlain by observer's concept that there exists a linear discriminant function $\mathbf{a}^T\mathbf{x}+b \gtrless 0$, namely, a discriminant hyperplane, determined by its direction vector in the feature space $\mathbf{a} \in \mathbb{R}^n$ and threshold $b \in \mathbb{R}$, such that primarily $\mathbf{a}^T\mathbf{x}+b>0$ if $y=1$ and <0 if $y=-1$. It is required to infer the hyperplane's parameters $(\hat{\mathbf{a}}, \hat{b})$ from a finite unordered training set $\{(\mathbf{x}_j, y_j), j=1, ..., N\}$ that contains information on both feature vectors and class-memberships of objects.

The problem of batch learning under concept drift differs from the classical statement of the pattern recognition problem in two aspects. First, the training set is no longer treated as completely unordered. Instead, a data stream of single training batches of individual size is considered $\{(\mathbf{x}_{j,t}, y_{j,t}), j=1, ..., N_t\}$, which arrive sequentially in time $t = 1, 2, 3, ...$ but are unordered each inside itself. Second, the observer's concept tolerates a relatively slow drift of the unknown discriminant hyperplane, i.e., its parameters may change in time $(\mathbf{a}_t, b_t)$.

We use the logistic regression approach [10], so, the probabilities of two possible class memberships of an instance $y_{j,t} = \pm 1$ can be expressed as a logistic functions of its feature vector $\mathbf{x}_{j,t}$

$$f(y_{j,t}|\mathbf{x}_{j,t}, \mathbf{a}_t, b_t) = \frac{1}{1+\exp\left[-(1/\sigma^2)y_{j,t}(\mathbf{a}_t^T\mathbf{x}_{j,t}+b_t)\right]}, \tag{1}$$

so that $f(1|\mathbf{x}_{j,t}, \mathbf{a}_t, b_t) + f(-1|\mathbf{x}_{j,t}, \mathbf{a}_t, b_t) = 1$. For the entire batch of training instances $\mathbf{X}_t = \{\mathbf{x}_{t,j}, j=1, ..., N_t\}$ and their class labels $\mathbf{Y}_t = \{y_{t,j}, j=1, ..., N_t\}$ the joint probability function is the product:

$$\Phi(\mathbf{Y}_t|\mathbf{X}_t, \mathbf{a}_t, b_t) = \prod_{j=1}^{N_t} f(y_{j,t}|\mathbf{x}_{j,t}, \mathbf{a}_t, b_t).$$

The key element of our Bayesian approach to the concept drift problem is treating the time-varying parameters of the hyperplane $(\mathbf{a}_t, b_t)$ as a hidden random processes that possesses the Markov property:

$$\begin{aligned} &\mathbf{a}_t = \mathbf{a}_{t-1} + \boldsymbol{\xi}_t, M(\boldsymbol{\xi}_t) = \mathbf{0}, M(\boldsymbol{\xi}_t\boldsymbol{\xi}_t^T) = d\mathbf{I}, \\ &b_t = b_{t-1} + \nu_t, M(\nu_t) = 0, M(\nu_t^2) = d'. \end{aligned} \tag{2}$$

Here $\boldsymbol{\xi}_t$ and ν_t are independent white noises with zero mathematical expectations, whose variances d and d' determine the assumed hidden dynamics of the concept.

The online learning protocol suggests that when the current training batch $(\mathbf{X}_t, \mathbf{Y}_t)$ comes, the observer retains only the last estimate of the discriminant hyperplane $(\hat{\mathbf{a}}_{t-1}, \hat{b}_{t-1})$ inferred from the precedent part of the data stream $(\mathbf{X}_s, \mathbf{Y}_s)_{s=1}^{t-1}$ that is no longer stored in the memory, and his/her task is to immediately recompute the current estimate with respect only to the new information $(\hat{\mathbf{a}}_t, \hat{b}_t) = \mathcal{F}\big[(\hat{\mathbf{a}}_{t-1}, \hat{b}_{t-1}), (\mathbf{X}_t, \mathbf{Y}_t)\big]$. In order to maximally save the statistical advantages of off-line learning $(\hat{\mathbf{a}}_t, \hat{b}_t) = \mathcal{F}\big[(\mathbf{X}_s, \mathbf{Y}_s)_{s=1}^{t}\big]$, let us temporarily assume that the entire prehistory of the data stream is still available.

The a priori distribution density of the hidden sequence of hyperplane parameters will have the form (2):

$$\Psi\Big((\mathbf{a}_s, b_s)_{s=2}^{t} | \mathbf{a}_1, b_1\Big) = \prod_{s=2}^{t} \psi_s(\mathbf{a}_s, b_s | \mathbf{a}_{s-1}, b_{s-1}),$$

$$\psi_s(\mathbf{a}_s, b_s | \mathbf{a}_{s-1}, b_{s-1}) \propto \mathcal{N}\big(\mathbf{a}_s | \underbrace{\mathbf{a}_{s-1}}_{\substack{\text{math.}\\ \text{exp.}}}, \underbrace{d\mathbf{I}}_{\substack{\text{covar.}\\ \text{matr.}}}\big)\mathcal{N}\big(b_s | \underbrace{b_{s-1}}_{\substack{\text{math.}\\ \text{exp.}}}, \underbrace{d'}_{\text{var.}}\big) =$$

$$\frac{1}{d^{n/2}(2\pi)^{n/2}} \exp\Big(-\frac{1}{2d}\big(\mathbf{a}_s - \mathbf{a}_{s-1}\big)^T\big(\mathbf{a}_s - \mathbf{a}_{s-1}\big)\Big) \times \frac{1}{2\pi d'} \exp\Big(-\frac{1}{2d'}(b_s - b_{s-1})^2\Big).$$

If we assume that there is no a priori information on the first value of the parameter vector $(\mathbf{a}_1, b_1)$, then the a posteriori distribution density of the entire hidden sequence of hyperplane parameters will be proportional to the product $P\big((\mathbf{a}_s, b_s)_{s=1}^{t} | (\mathbf{X}_t, \mathbf{Y}_s)_{s=1}^{t}\big) \propto \Phi\big((\mathbf{Y}_s)_{s=1}^{t} | (\mathbf{X}_s, \mathbf{a}_s, b_s)_{s=1}^{t}\big)\Psi\big((\mathbf{a}_s, b_s)_{s=1}^{t}\big)$.

The sought-for estimate of time-varying parameters $(\hat{\mathbf{a}}_s, \hat{b}_s)_{s=1}^{t}$ is the maximum point of the joint distribution of the parameters and the training set:

$$(\hat{\mathbf{a}}_s, \hat{b}_s)_{s=1}^{t} = \arg\max_{\mathbf{a}_s, b_s} P\big((\mathbf{a}_s, b_s)_{s=1}^{t} | (\mathbf{X}_s, \mathbf{Y}_s)_{s=1}^{t}\big) = \arg\min_{\mathbf{z}_s} J_t\big((\mathbf{z}_s)_{s=1}^{t}\big), \qquad (3)$$

$$J_t\big((\mathbf{z}_s)_{s=1}^{t}\big) = \sum_{s=1}^{t} \underbrace{\sum_{j=1}^{N_s} \ln\big(1 + \exp(-C\mathbf{g}_{j,s}^T \mathbf{z}_s)\big)}_{\eta_s(\mathbf{z}_s)} +$$

$$+ \sum_{s=2}^{t} \underbrace{(\mathbf{z}_s - \mathbf{z}_{s-1})^T \mathbf{U} (\mathbf{z}_s - \mathbf{z}_{s-1})}_{\gamma_s(\mathbf{z}_{s-1}, \mathbf{z}_s)} = \sum_{s=1}^{t} \eta_s(\mathbf{z}_s) + \sum_{s=2}^{t} \gamma_s(\mathbf{z}_{s-1}, \mathbf{z}_s) \to \min_{\mathbf{z}_s}, \qquad (4)$$

$$\mathbf{z}_s = \begin{bmatrix} \mathbf{a}_s \\ b_s \end{bmatrix} \in \mathbb{R}^{n+1}, \ \mathbf{g}_{j,s} = \begin{bmatrix} y_{j,s}\mathbf{x}_{j,s} \\ y_{j,s} \end{bmatrix} \in \mathbb{R}^{n+1}, \ C = \frac{2}{\sigma^2}, \ \mathbf{U} = Diag\Big\{\frac{1}{d} \cdots \frac{1}{d}, \frac{1}{d'}\Big\}.$$

The criterion (4) is pair-wise separable, i.e., has the structure of a sum of functions each of which is associated with one time point (s) or two immediately adjacent time points in their increasing order $(s-1, s)$. This computational problem is, generally speaking, that of dynamic programming.

In the online learning mode, only the last element $\hat{\mathbf{z}}_t = (\hat{\mathbf{a}}_t, \hat{b}_t)$ of the estimated sequence $(\hat{\mathbf{z}}_s)_{s=1}^{t}$ is of interest at each current time point t. The main notion of dynamic programming is that of the sequence of Bellman functions $\tilde{J}_t(\mathbf{z}_t)$ whose minimum points yield just the required online estimates:

$$\tilde{J}_t(\mathbf{z}_t) = \min_{\mathbf{z}_1,\ldots,\mathbf{z}_{t-1}} J_t\Big((\mathbf{z}_s)_{s=1}^t\Big), \quad \hat{\mathbf{z}}_t = \arg\min_{\mathbf{z}_t} \tilde{J}_t(\mathbf{z}_t). \tag{5}$$

The fundamental property of Bellman functions is the almost evident equality

$$\tilde{J}_t(\mathbf{z}_t) = \eta_t(\mathbf{z}_t) + \min_{\mathbf{z}_{t-1}} \Big[\gamma_t(\mathbf{z}_{t-1}, \mathbf{z}_t) + \tilde{J}_{t-1}(\mathbf{z}_{t-1})\Big], \tag{6}$$

which would immediately suggest an algorithm of their sequential computation if only there existed a simple way of solving the optimization problem in (6) with real-valued vectors $\mathbf{z}_{t-1}$.

If all the items of the pair-wise separable learning criterion (4) were quadratic, each of the Bellman functions (5)-(6) would be also quadratic

$$\tilde{J}_t(\mathbf{z}_t) = (\mathbf{z}_t - \tilde{\mathbf{z}}_t)^T \tilde{\mathbf{Q}}_t (\mathbf{z}_t - \tilde{\mathbf{z}}_t) + \tilde{c}_t. \tag{7}$$

and easily computable in terms of their parameters $\big[\tilde{\mathbf{z}}_t \in \mathbb{R}^{n+1}, \tilde{\mathbf{Q}}_t\big((n{+}1){\times}(n{+}1)\big), \tilde{c}_t \in \mathbb{R}\big]$ by the classical Kalman-Bucy filter [11,12]. But only the second sum in (4) is quadratic, whereas the first one is formed by logarithmic items. Therefore, an approximation of Bellman functions is required before immediate application of the exceptionally effective Kalman-Bucy filter.

3 Approximate Dynamic Programming for Online Estimating the Time-Varying Discriminant Hyperplane

A method of overcoming the obstacle of non-quadratic Bellman functions (5)-(6) is proposed in [13]. However, that way is based on the assumption that the size of the training set t is much greater than the number of features n, and that both classes are equally represented in the training set. The former of these assumptions is simply satisfiable in the case of a data stream, but the latter one can be mistaken in very many cases. We suggest here a more universal approach to quadratic approximation of logarithmic summands in (4).

Our procedure is based on the assumption that there exists an approximate compact representation of Bellman fuctions, which permits storing them in the memory. The previous Bellman function $\tilde{J}_{t-1}(\mathbf{z}_{t-1})$ in (6) is non-quadratic because of the logarithmic term in (4). The idea of approximate implementation of the dynamic programming procedure consists in the substitution of the function

$$F_t(\mathbf{z}_t) = \min_{\mathbf{z}_{t-1}} \Big[\gamma_t(\mathbf{z}_{t-1}, \mathbf{z}_t) + \tilde{J}_{t-1}(\mathbf{z}_{t-1})\Big] \tag{8}$$

by an appropriate quadratic function

$$\bar{F}_t(\mathbf{z}_t) = (\mathbf{z}_t - \bar{\mathbf{z}}_t)^T \bar{\mathbf{Q}}_t (\mathbf{z}_t - \bar{\mathbf{z}}_t), \tag{9}$$

then the following approximations of the next Bellman function $\tilde{J}_t(\mathbf{z}_t)$ will be quadratic too. It remains to choose appropriate values of the parameters $(\bar{\mathbf{z}}_t, \bar{\mathbf{Q}}_t)$ of the quadratic function $\bar{F}_t(\mathbf{z}_t)$, which would ensure conservation of the main aspects of, generally speaking, non-quadratic original Bellman function. We propose to retain the minimum point of the function $\bar{\mathbf{z}}_t = \arg\min F_t(\mathbf{z_t})$ and the Hessian at the minimum point $\bar{\mathbf{Q}}_t = \nabla^2 F_t(\bar{\mathbf{z}}_t)$.

Theorem 1. *Let function* (8) *have the form*

$$F_t(\mathbf{z}_t) = \min_{\mathbf{z}_{t-1}} \left[(\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1})^T \mathbf{U} (\mathbf{z}_t - \mathbf{A}\mathbf{z}_{t-1}) + \sum_{j=1}^{N_{t-1}} \ln\left(1 + \exp(-C\mathbf{g}_j^T \mathbf{z}_{t-1})\right) \right].$$

Then the parameters $(\bar{\mathbf{z}}_\mathbf{t}, \bar{\mathbf{Q}}_\mathbf{t})$ *of approximation* (9) *are determined as*

$$\bar{\mathbf{Q}}_t = \tilde{\mathbf{Q}}_{t-1} \left(\mathbf{AUA} + \tilde{\mathbf{Q}}_{t-1} \right)^{-1} \mathbf{U}, \; \bar{\mathbf{z}}_t = \bar{\mathbf{Q}}_t \left[\left(\mathbf{AUA} + \tilde{\mathbf{Q}}_{t-1} \right)^{-1} \mathbf{AU} \tilde{\mathbf{Q}}_{t-1} \tilde{\mathbf{z}}_{t-1} \right].$$

4 Experimental Results

In order to test our algorithm, we got the Luxembourg dataset by Zliobaite [16], which is constructed using the European Social Survey on internet usage [15]. The dataset is set of 1901 answers to the survey questionnaire, each represented by 31 real-valued features transformed from the original categorical variables in the survey format, and supplied by a time stamp over a 5 year period. The task is to classify the instances with respect to the internet usage high or low under the assumption of concept drift, i.e., internet usage changing in time.

For the experiment, we took the data part from the 851th instance to 1901th one, within which the time increment is relatively constant, because our a priori model of concept drift (2) does not observe time stamps. The entire time series was separated its into 105 batches of 10 instances each.

In the process of online classifier evaluation, before training on the next batch, we computed the error rate on it with the current decision rule. The total error was calculated as the average value of error rates in all the successive batches.

Table 1 shows the comparative results of our algorithm abbreviated as **LogitClassifierDrift** with two algorithms of concept drift realized in the software environment Massive Online Analysis (MOA), see [17] for detail.

Table 1. The experimental results: Luxembourg dataset

Algorithm	Classification error, %
OzaBagAdwin (16 models)	14.519
SingleClassifierDrift	16.36
LogitClassifierDrift	14.381

5 Conclusions

In this paper we present the mechanism for the problem under concept drift with data streams. This mechanism is based on the Bayesian approach to the logistic regression in the finite-dimensional feature space under Markov assumption on the drifting discriminant hyperplane. The error rate indicates that the proposed algorithm outperforms the benchmark ones in the dataset of the European Social Survey on internet usage.

References

1. Polikar, R., Upda, L., Upda, S.S., Honavar, V.: Learn++: an incremental learning algorithm for supervised neural networks. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 31(4), 497–508 (2001)
2. Mongillo, G., Deneve, S.: Online learning with hidden Markov models. Neural Computation 20, 1706–1716 (2008)
3. Florez-Larrahondo, G., Bridges, S., Hansen, E.A.: Incremental estimation of discrete hidden Markov models based on a new backward procedure. In: Proceedings of the 20th National Conference on Artificial intelligence, vol. 2, pp. 758–763 (2005)
4. Ulas, A., Semerci, M., Yildiz, O.T., Alpaydin, E.: Incremental construction of classifier and discriminant ensembles. Information Sciences 179, 1298–1318 (2009)
5. Kapp, M., Sabourin, R.R., Maupin, P.: Adaptive incremental learning with an ensemble of support vector machines. In: Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, pp. 4048–4051 (2010)
6. Elwell, R., Polikar, R.: Incremental Learning of Concept Drift in Nonstationary Environments. IEEE Transactions on Neural Networks 22(10), 1517–1531 (2011)
7. Widmer, G., Kubat, M.: Learning in the presence of concept drift and hidden contexts. Machine Learning 23, 69–101 (1996)
8. Bifet, A., Gavalda, R.: Learning from time-changing data with adaptive windowing. In: SIAM International Conference on Data Mining (2007)
9. Wang, H., Fan, W., Yu, P.S., Han, J.: Mining concept-drifting data streams using ensemble classifiers. In: Proceedings of the Ninth ACM SIGKDD International Conference KDD 2003, pp. 226–235. ACM Press (2003)
10. Bishop, C.M.: Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag New York, Inc., Secaucus (2006)
11. Markov, M., Krasotkina, O., Mottl, V.: Time-varying regression model with unknown time-volatility for nonstationary signal analysis. In: Proceedings of the 8th IASTED International Conference on Signal and Image Processing, Honolulu, Hawaii, August 14-16, pp. 534–196 (2006)
12. Grewal, M.S., Andrews, A.P.: Kalman Filtering: Theory and Practice Using MATLAB. Wiley (2008)
13. Turkov, P., Krasotkina, O., Mottl, V.: Bayesian Approach to the Concept Drift in the Pattern Recognition Problems. In: Perner, P. (ed.) MLDM 2012. LNCS (LNAI), vol. 7376, pp. 1–10. Springer, Heidelberg (2012)
14. Street, W., Kim, Y.: A streaming ensemble algorithm (SEA) for large-scale classification. In: Proc. 7th Int. Conf. on Knowledge Discovery and Data Mining KDD 2001, pp. 377–382 (2001)
15. Jowell, R., The Central Coordinating Team.: European social survey 2002/2003; 2004/2005; 2006/2007. Technical Reports, London: Centre for Comparative Social Surveys, City University (2003, 2005, 2007)
16. Žliobaité, I.: Combining time and space similarity for small size learning under concept drift. In: Rauch, J., Raś, Z.W., Berka, P., Elomaa, T. (eds.) ISMIS 2009. LNCS (LNAI), vol. 5722, pp. 412–421. Springer, Heidelberg (2009)
17. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B.: MOA: Massive Online Analysis. Journal of Machine Learning Research 11, 1601–1604 (2010), `http://sourceforge.net/projects/moa-datastream/`

Gait Recognition from Front and Back View Sequences Captured Using Kinect

Pratik Chattopadhyay[1], Shamik Sural[1], and Jayanta Mukherjee[2]

[1] School of Information Technology, IIT Kharagpur, India
[2] Dept. of Computer Science & Engineering, IIT Kharagpur, India
{pratikc,shamik@sit.iitkgp.ernet.in,
jay@cse.iitkgp.ernet.in

Abstract. In this paper, we propose a key pose based gait recognition approach using skeleton joint information derived from the depth data of Kinect. We consider situations where such depth cameras are mounted on top of entry and exit points, respectively capturing back and front views of subjects who enter a zone under surveillance. Three dimensional geometric transformations are used to map the skeleton images captured from the back view to an equivalent front view. A gait cycle is divided into a number of key poses and the trajectory followed by each skeleton joint within a key pose is used to derive the gait features for that particular pose. For recognizing a subject, available key poses are compared with the corresponding key poses of the training subjects. The proposed method has higher accuracy than other competing approaches.

Keywords: Gait Recognition, Kinect, Geometric Transformation, Skeleton Joint, Key Pose.

1 Introduction and Background

Gait recognition is the process of identifying an individual from his style of walking. Suitability of gait as a biometric feature in security applications is established by the fact that unlike other biometric methods like face detection, iris scan and finger print detection, close interaction with a subject is not required in gait recognition. Gait can be identified from a distance and even with low resolution video sequences.

Most of the existing work on gait recognition considers capturing of video sequences in the fronto-parallel (side) view, since maximum gait information during walking is contained in this view [1], [2]. It has, however, been pointed out that the assumption on the availability of gait features from the fronto-parallel view is too restrictive especially in situations where gait is presumed to be the biometric of choice, i.e., in places like airports and railway stations [3]. In such places, people are often directed to enter into secured zones one at a time like walking through a metal detector gate. Since it is easier to control the flow of persons at the entrance/exit of a security zone, surveillance cameras mounted above the entry/exit point can record a better view of a subject entering/leaving

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 196–203, 2013.

the zone. Such type of camera positioning captures the frontal view (back/front view) of an individual.

Research work strictly based on frontal gait recognition includes the curve spread method [4], methods based on intra and inter frame rectangular size distribution [5], the largest Lyapunov exponents [6], three dimensional moments from silhouette shape variations [7] and a spherical space model for gait recognition [8]. These frontal gait recognition procedures have worked with gait videos captured by RGB video camera, and have shown accuracy to a certain extent. Frontal gait recognition based on silhouette shape and depth has started after RGB-D cameras like Kinect were made publicly available [9]. Partial volumetric reconstruction of the depth data provided by Kinect has been used to construct Gait Energy Volumes (GEV) [3]. To preserve the dynamic information of gait at a higher resolution than what was considered in [3], Pose Depth Volume (PDV) [10] has been recently developed. These methods are computationally intensive as they operate at the voxel level of granularity. A comparatively faster gait recognition procedure based on the covariance of the skeleton joint trajectories provided by Kinect [11] has been proposed in [12].

The present study also makes use of the skeleton joint positions but the kinematic information of gait is preserved at a higher resolution by dividing a gait cycle into a number of key poses and constructing a gait feature corresponding to each such key pose.

The organization of the rest of the paper is as follows. Section 2 presents the proposed method for estimation of key poses, whereas construction of gait features and human recognition using the proposed feature are detailed in Section 3. The experimental setup together with results is presented in Section 4. Concluding remarks and future scope of work are pointed out in Section 5.

2 Key Pose Estimation

Gait is a repetitive pattern followed during walking where each gait cycle covers two strides: the right foot forward and the left foot forward. Key poses are representatives of an entire gait cycle that characterize the gait cycle by maintaining sufficient variability between any two consecutive key poses. In this section, we describe the steps used in key pose estimation.

2.1 Skeleton Alignment and Feature Vector Construction

Let the angle of inclination of the skeleton with respect to the Y axis of the Kinect coordinate system be α. This inclination angle always causes the skeleton to be tilted at an angle with respect to the YZ plane of the Kinect coordinate system. But for frontal view gait videos, any inclination with respect to the XY plane of the coordinate system can be ignored. To make key pose estimation and gait recognition procedures invariant to camera positioning, transformation of the skeleton to a uniform viewpoint is accomplished using a three dimensional geometric transformation. In the present context, the segment joining the torso

to neck of each skeleton is aligned along the XY plane of the Kinect coordinate system. Let the 3D coordinates of the torso and the neck of a certain skeleton as provided by Kinect be $T(X_t, Y_t, Z_t)$ and $N(X_n, Y_n, Z_n)$, respectively. First, each joint of the skeleton is translated by an amount $(-X_t, -Y_t, -Z_t)$, so that the torso of the skeleton maps to the origin. The translated coordinates of the torso and the neck are now given by $T'(0,0,0)$ and $N'(X'_n, Y'_n, Z'_n)$ respectively. Next, the skeleton is rotated about Z axis by an angle α so that the rotated coordinates of the neck maps to the XY plane of the coordinate system. For any transformation along the YZ plane, the value of the X coordinate does not alter. Thus, if $N''(X'_n, Y''_n, 0)$ denotes the new transformed coordinates of the neck, the numerical value of α is given by:

$$\alpha = cos^{-1}(Y''_n/L), \tag{1}$$

where, L is the length of the segment joint N' with the origin.

For determining the key poses, consider a set $S = \{S_1, S_2, ..., S_N\}$ of N skeletons comprising of walking stances of a sufficiently large number of persons. These skeletons are available either from the front view or from the back view. Consider that C skeleton joints are tracked for each subject. For each skeleton, we compute the direction cosines of each of its joints with respect to the torso (i.e., origin). For any two vectors $\overrightarrow{A}$ and $\overrightarrow{B}$, the unit vector pointing from $\overrightarrow{A}$ to $\overrightarrow{B}$ denoted by $\widehat{AB}$ is given by:

$$\widehat{AB} = l\hat{i} + m\hat{j} + n\hat{k}, \tag{2}$$

where, l, m and n are termed as the direction cosines of the vector $\overrightarrow{AB}$. Let l_c, m_c and n_c denote the direction cosines of the c^{th} joint with respect to the torso, where, c = 1, 2, 3, ..., C. Let V_i denote the feature vector obtained after concatenating the direction cosines of all the C - 1 joints with respect to the torso for the i^{th} skeleton. Thus, V_i is represented as:

$$V_i = [l_c^1 \;\; m_c^1 \;\; n_c^1 \;\; l_c^2 \;\; m_c^2 \;\; n_c^2 \;\; ... \;\; l_c^{C-1} \;\; m_c^{C-1} \;\; n_c^{C-1}]^T, \tag{3}$$

for i = 1, 2, ..., N. Similar feature vectors computed for all the N skeletons in the training set are concatenated to form the feature set for key pose estimation. Since the skeleton joints must be selected in a specific order, during feature construction, if the skeleton joints are chosen in clockwise order for front view skeletons, the ordering of the same sequence of joints for back view skeletons will be anticlockwise. Again, for each skeleton obtained from the back view, the signs of the X and Z components of the direction cosines will be reversed when the front view direction cosines have to be computed from the back view.

2.2 Key Pose Estimation by Feature Vector Clustering

Constrained K-Means clustering [13] is applied on the feature vectors constructed from the skeletons to determine the key poses. The centroids of the clusters obtained after clustering are the key poses, which we term as *skeleton*

key poses. The constraint imposed on the cluster assignment step of the clustering process is that, reassignment of any feature vector in a subsequent step is only valid for the immediately preceding or succeeding cluster indices. The skeleton key poses are labeled such that key pose *1* starts with the mid stance and key pose K ends with the pose just before the mid stance.

3 Gait Feature Extraction and Human Recognition

For deriving the gait features, the transformed joint positions on each skeleton rather than the direction cosines relative to the torso are considered. This is because the transformed joint position coordinates preserve information regarding local motion characteristics at a higher resolution than the direction cosines. Since the torso is always fixed at origin, the mean position of each of the remaining C - 1 joints over all the skeletons belonging to a particular key pose is used to form the feature vector for that key pose.

3.1 Feature Extraction

For an input training gait sequence of a subject, we consider unconstrained data where the front view or the back view sequence could be available. Three dimensional transformations as explained in Section 2 are applied on each skeleton and equivalent front view representation is derived for each skeleton from the back view.

To find the mapping between the key pose numbers and the skeleton sequence, dynamic programming based Viterbi path decoding algorithm is employed. This step assigns to each skeleton of the sequence a unique key pose number. The assignment preserves a cyclic ordering of natural numbers 1 to K. For example, if the i^{th} skeleton is assigned a key pose k, then the $(i+1)^{th}$ skeleton must be assigned to either key pose k or key pose $1 + (k \bmod K)$. Let the sequence of p skeletons $P = \{S_n, S_{n+1}, S_{n+2}, ..., S_{n+p-1}\}$ (which is a subset of the input sequence $Q = \{S_1, S_2, S_3, ..., S_Q\}$) of a certain subject map to the key pose k, k = 1, 2, 3, ..., K. Further, let $J_{c,r}(X_c^r, Y_c^r, Z_c^r)$ be the transformed coordinates of the c^{th} joint of the r^{th} skeleton (S_r), that is mapped to the key pose k, where, $r = n$, n+1, n+2, ..., n+p and c = 1, 2, 3, ..., C - 1. The gait feature vector F_k corresponding to the k^{th} skeleton key pose is hence obtained by averaging the X, Y and Z coordinates of each of the C - 1 joints of the skeletons that are mapped to the key pose k. Mathematically,

$$F_k = [X_{1k}\ \ Y_{1k}\ \ Z_{1k}\ \ X_{2k}\ \ Y_{2k}\ \ Z_{2k}\ \ X_{3k}\ \ Y_{3k}\ \ Z_{3k}\ \ ...\ \ X_{(C-1)k}\ \ Y_{(C-1)k}\ \ Z_{(C-1)k}]^T. \tag{4}$$

where,

$$X_{ck} = \frac{1}{p}\sum_{r=n}^{n+p-1} X_{rk}, \qquad Y_{ck} = \frac{1}{p}\sum_{r=n}^{n+p-1} Y_{rk}, \qquad Z_{ck} = \frac{1}{p}\sum_{r=n}^{n+p-1} Z_{rk}, \tag{5}$$

Similar feature vectors are constructed for all the K key poses and are concatenated to form the training gait feature set. Thus, the k^{th} column of the training gait feature vector represents the gait feature of the subject for the k^{th} key pose.

Now, consider an input test sequence which could be obtained from the back view only, from the front view only, or from each of the two views. The transformed skeleton sequences recorded from the front view and/ or the back view are taken into consideration separately. Viterbi path decoder is applied on each of these two sequences in order to obtain the skeleton key poses of each skeleton in the sequence.

Construction of the test gait feature vector for a particular key pose follows a similar procedure as that during training, but the computation considers those skeletons that are available either from the front or the back view or from both the views.

3.2 Recognition Using Proposed Feature

Whichever frames of the partial gait cycles are obtained from the two cameras, they are mapped to the appropriate skeleton key poses and searching within the training set is restricted to these skeleton key poses only.

Let the training gait features be available for $\mathcal{S}$ subjects. Let F_1^i, F_2^i, F_3^i, ..., F_K^i be the K feature vectors corresponding to the K skeleton key poses of the i^{th} subject, where, $i = 1, 2, 3, ..., \mathcal{S}$. Suppose $F_j^i = [F_{1j}^i \; F_{2j}^i \; F_{3j}^i \; ... \; F_{\mathcal{N}j}^i]^T$, where $\mathcal{N} = 3 \times (C - 1)$ and j ranges from 1 to K. Suppose for an input test subject, walking stances belonging to only l out of the K skeleton key poses are available. It may be noted that certain key poses could be missing even after combining front and back view sequences due to occlusion or noisy video capture. Let k_1, k_2, k_3, ..., k_l be the available skeleton key poses for the test subject, where, $1 \leq k_m \leq K$, $\forall \; m = 1, 2, ..., l$. Also, let us suppose that F'_{k_1}, F'_{k_2}, F'_{k_3}, ..., F'_{k_l} be the respective gait features for the available skeleton key poses of the test subject computed in a manner similar to that explained in Section 3. The gait feature vectors corresponding to these available skeleton key poses of the test subject are compared with the feature vectors corresponding to the same skeleton key poses of the training subjects using Euclidean distance as the metric for comparison. The training subject with the minimum value of the Euclidean distance is assigned to be the class of the test subject. Thus, for each of the $\mathcal{S}$ subjects, we compute the following:

$$D_i = \sum_{m=1}^{l} \sum_{p=1}^{\mathcal{N}} (F_{pk_m}^i - F'_{pk_m})^2. \tag{6}$$

The test subject is assigned to the r^{th} class, if $D_r < D_i$, $\forall$ i = 1, 2, 3, ..., $\mathcal{S}$ and $i \neq r$.

4 Experimental Results

We next present results from an extensive set of experiments carried out using the proposed method. The algorithm has been implemented in Matlab 7.12.0 (R2010a) on a 2.50 GHz Intel Core i5 processor having 4GB RAM. The experimental setup is described next.

Two Kinects are positioned at two ends of a narrow pathway, facing each other, each at a height of 2 metres from the ground. The tilt angles of both the Kinects are set to -27 degrees. During both the key pose estimation phase and the gait training phase, only one Kinect camera is employed to record the gait sequence of each subject from the front view as well as the back view. The test set is constructed by allowing each person to walk through the pathway. The data set consists of a total of 45 subjects. Two complete gait cycles *T1* and *T2* from the front and two complete cycles *T3* and *T4* from the back, each captured at 30 frames per second (fps), serve as the training data set. The entire video recording procedure for the test cases is repeated for two different walking speeds: normal walk (NW), i.e., at the same speed as that of the training sequence and fast walk (FW), both of which are again recorded under two different frame rate settings: *F1* = 30 fps and *F2* = 15 fps, respectively. For comparative analysis with other recently proposed gait features, we have captured the RGB as well as depth channel information along with skeleton joint coordinates for a total of 17 subjects, selected out of these 45 subjects.

First, we study the performance of the proposed method in case of both normal and fast walk. Results are presented in Table 1 with two different frame rates. The different rows represent various combinations of data sets (*T1*, *T2*, *T3*, *T4*) used for training as mentioned in the last sub-section. It is observed that quite accurate results are obtained both for normal walk and fast walk when the frame rate during test data capture is kept the same as that during the training data capture. A higher performance is usually achieved with normal walk than with fast walk because fast walk causes greater number of key poses to be missed in a walking sequence. Also, not much degradation is observed by reducing the frame rate to half in the test case. In general, performance of the algorithm improves if the size of the training data is increased.

In most practical situations, we do not need to choose only the subject with the best match. Rather, it is sufficient if the correct class falls within the top r classes predicted by the algorithm. In the present study, we examine how many subjects are correctly classified as the value of r is varied from 1 to 10. The training set for this experiment consists of *T1* and *T3* (i.e., one gait cycle each from the front and back views) and the results are shown using the cumulative match characteristic (CMC) curves of Figure 1. It is observed that beyond a rank of 5, almost 98% accuracy is obtained. Further, similar performance is achieved for both fast walk and slow walk.

Finally, a comparative analysis of performance of the proposed method is made with some of the existing gait recognition approaches, namely, those derived by utilizing the depth data of Kinect like Gait Energy Volume [3] and skeleton co-variance feature [12] along with the widely used Gait Energy Image feature [1]. Results are presented in Table 2. Training set for this experiment consists of *T1* and *T3*. It is seen that although the proposed method takes about 1 second more than each of [1] and [12], it outperforms all of the current state of the art algorithms in terms of accuracy by more than 5%, which is quite significant.

Table 1. Variation of recognition accuracy (in %) with frame rate (*F1* and *F2*) for both normal walk (*NW*) and fast walk (*FW*) with different training data set (*T1*, *T2*, *T3* and *T4*) combinations

Training Data Set	*F1* (frame rate: 30 fps)		*F2* (frame rate: 15 fps)	
	NW	*FW*	*NW*	*FW*
T1	80.00	77.78	82.22	66.67
T2	75.56	84.44	84.44	62.22
T3	80.00	84.44	80.00	68.89
T4	80.00	73.33	68.88	71.11
T1+T2	82.22	82.22	91.11	73.33
T1+T3	93.33	93.33	93.33	84.44
T2+T4	93.33	91.11	91.11	77.78
T3+T4	86.67	80.00	84.44	84.44
T1+T2+T3+T4	95.56	95.56	95.56	88.89

Table 2. Performance comparison of different gait recognition algorithms

Gait Feature	Accuracy(%)	Time(secs)
Gait Energy Image [1]	76.47	0.45
Gait Energy Volume [3]	70.59	2.33
Skeleton Co-Variance Feature [12]	76.47	0.30
Proposed feature	82.35	1.37

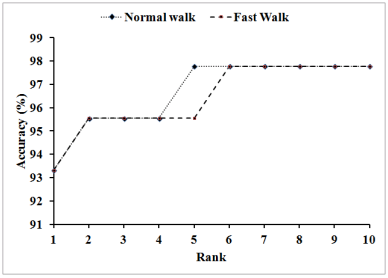


Fig. 1. Cumulative match characteristic curves with two different walking speeds during testing

5 Conclusion

In this paper, we have proposed a key pose based approach of gait recognition. The gait features for each key pose are derived by utilizing the skeleton tracking data provided by Kinect. Better skeleton tracking capability of Kinect is expected to improve the performance of the algorithm even further. From the extensive set of experiments, it can be concluded that use of Kinect depth data has an immense potential in frontal gait recognition. The method is also fast enough to achieve meaningful response time. The processing time can be reduced by using coding in C/ C++ instead of Matlab. More experiments, however, need to be conducted with challenging conditions including occlusion, collision, etc.

Acknowledgement. This work is partially funded by Project Grant No. 22(0554)/11/EMR-II sponsored by the Council of Scientific and Industrial Research, Govt. of India.

References

1. Han, J., Bhanu, B.: Individual Recognition using Gait Energy Image. IEEE Trans. on Pattern Analysis and Machine Intel. 28(2), 316–322 (2006)
2. Roy, A., et al.: Gait Recognition using Pose Kinematics and Pose Energy Image. Signal Processing 92(3), 780–792 (2012)
3. Sivapalan, S., et al.: Gait Energy Volumes and Frontal Gait Recognition using Depth Images. In: International Joint Conf. on Biometrics, pp. 1–6 (2011)
4. Soriano, M., et al.: Curve Spreads - A Biometric From Front View Gait Video. Pattern Recognition Letters 25(14), 1595–1602 (2004)
5. Barnich, O., et al.: Frontal-View Gait Recognition by Intra- and Inter-Frame Rectangle Size Distribution. Pattern Recognition Letters 30(10), 893–901 (2009)
6. Lee, K., et al.: Frontal View-Based Gait Identification using Largest Lyapunov Exponents. In: Proc. of the Intl. Conf. on Accoustics, Speech, and Signal Processing, pp. 173–176 (2006)
7. Gofferedo, M., et al.: Front-View Gait Recognition. In: Proc. of the IEEE Intl. Conf. on Biometrics: Theory, Applications and Systems, pp. 1–6 (2011)
8. Ryu, J., Kamata, S.: Front View Gait Recognition using Spherical Space Model with Human Point Clouds. In: Eighteenth IEEE International Conf. on Image Processing, pp. 3209–3212 (2011)
9. Zhang, Z.: Microsoft Kinect Sensor and its Effect. IEEE Multimedia 19(2), 4–10 (2012)
10. Chattopadhyay, P., et al.: Pose Depth Volume Extraction from RGB-D Streams for Frontal Gait Recognition. Journal of Visual Communication and Image Representation (2013), doi:10.1016/j.jvcir.2013.02.010
11. Shotton, J., et al.: Real-Time Human Pose Recognition in Parts from Single Depth Images. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1297–1304 (2011)
12. Kumar, M.S., Babu, R.V.: Human Gait Recognition using Depth Camera: A Covariance Based Approach. In: Proc. of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, Art., vol. 20 (2012)
13. Jin, X., Han, J.: K-Means Clustering. In: Encyclopedia of Machine Learning, pp. 563–564. Springer, US (2010)

Unconstrained Kannada Handwritten Character Recognition Using Multi-level SVM Classifier

G.G. Rajput and Rajeshwari Horakeri

Department of Computer Science,
Gulbarga University, Gulbarga
Karnataka State, India
{ggrajput,rajeshwari_horakeri}@yahoo.co.in

Abstract. This paper presents an efficient zoning based method for recognition of handwritten Kannada characters using two sets of features, namely, crack codes (the line between the object pixel and background) and the density of the object pixels. A multi-level SVM is used for the classification purpose. The proposed method is implemented in two stages. In the first stage, similar shaped characters are combined into groups resulting in 22 classes instead of 49 classes, one class per character. Crack codes are used to assign the input character image to one of the groups. In the second stage, object pixel density is used to assign label to the input character image within that identified group. Experiments are performed on handwritten Kannada characters consisting of 24500 images with 500 samples for each character. Five-fold cross validation is used for result computation and average recognition rate of 91.02 % is obtained.

Keywords: Kannada, Crack codes, SVM, handwritten character, five-fold cross validation.

1 Introduction

Handwritten Character Recognition (HCR) is one of the important areas in pattern recognition field because it provides solution for document classifications, postal automation, bank check reading, automatic data entry, reading customer filled forms, etc. As a result of intensive research, several systems are available in many non-Indian scripts like Chinese, Arabic and English [1,2,3]. HCR is a challenging task as compared to recognition of printed characters due to unconstrained handwriting, touching of adjacent characters, broken strokes, poor binarization and noise. The task is more complicated for Indian languages due to complexity in the shape and number of characters compared to characters of non-Indian scripts like English and Chinese. A brief review of the work done towards the recognition of characters written in Indian scripts is presented below.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 204–212, 2013.

1.1 Related Works

A review on recognition of handwritten characters of south Indian scripts is presented in [4] and [5]. Printed and handwritten character recognition of Devanagari script using gradient features is presented in [6]. Gradient features are extracted using Sobel and Robert operators and fed to SVM classifier for recognition. Recognition of off-line Bangla handwritten compound characters using modified quadratic discriminant function (MQDF) is reported in [7]. Features used for recognition purpose are based on directional information obtained from the arc tangent of the gradient. A Robertss filter is applied on the normalized image to obtain gradient image. Next, the arc tangent of the gradient is initially quantized into 32 directions and the strength of the gradient is accumulated with each of the quantized direction. Finally, the frequencies of these directions are down sampled using Gaussian filter to obtain 392 dimensional feature vectors.

Recognition of handwritten Bangla basic characters is reported in [8]. The features are obtained by computing local chain code histograms of input image and the classification results are obtained after down sampling the histogram feature by applying Gaussian filter. Multilayer perceptrons (MLP) trained by back propagation (BP) algorithm is used for classification. Fuzzy model based recognition of handwritten Hindi characters is presented in [9]. Features consists of normalized distances obtained using the box approach. Pal et. al. [10] have proposed a quadratic classifier based scheme for the recognition of off-line handwritten characters of three popular south Indian scripts, namely, Kannada, Telugu, and Tamil. The bounding box of a character is segmented into blocks, and the directional features are computed in each block. These blocks are then down-sampled by a Gaussian filter, and the features so obtained are fed to a modified quadratic classifier for recognition. The mean accuracy of 90.34%, is reported for Kannada characters. FLD based unconstrained handwritten Kannada characters recognition is presented in [11]. The mean recognition accuracy of 68% is reported using Euclidean distance measure. In [12], handwritten Kannada character recognition system based on spatial features viz stroke density, stroke length and the number of stokes is proposed. Using k-NN classifier, the average recognition accuracy of 90.1% is reported for vowel characters. In [13], moment based features are used for recognition of Kagunita (the Kannada compound characters resulting from the consonant and vowel combination). These features are extracted using Gabor wavelets from the dynamically preprocessed original image. Multi-Layer Perceptron with Back Propagation Neural Networks are employed for character classification. Average recognition rate of 86% is reported for vowels and for consonants the average recognition reported is 65%. In [14], features are extracted using Fourier Transform and PCA (Principal Component Analysis) and recognition accuracy of 68.89% is obtained using PNN. A Neural network based Kannada character recognition system with limited set of characters with an accuracy of 99.17 % is presented in [15].

From the literature survey, it is clear that very few attempts have been made for handwritten Kannada characters with limited number of characters taken from the alphabet set (vowels and consonants). Further, the dataset used for

performing experiments is limited to small number of samples as compared to number of samples available in standard dataset of other languages like English. This motivated us to propose an efficient method for handwritten Kannada character recognition by including the entire alphabet set and creating a large dataset for performing experiments. This is challenging task as it results in more number of classes to be labeled.

The rest of the paper is organized as follows. In Section 2, handwritten Kannada characters database creation is presented. Section 3 deals with the feature extraction and section 4 describe classification. The experimental results are presented in Section 5 and conclusion is given in section 6.

2 Database Creation

Kannada, the official language of the southern Indian state of Karnataka, is a Dravidian language. The Kannada alphabet was developed from the Kadamba and Chalukya scripts used between the 5th and 7th centuries AD. Under the influence of Christian missionary organizations, Kannada script was standardized at the beginning of the 19th century. The writing style is from left to right. Kannada is built up from a base character set of 49 characters: 15 vowels and 34 consonants. Fig. 1 present a listing of the characters (vowels and consonants) used in Kannada scripts.

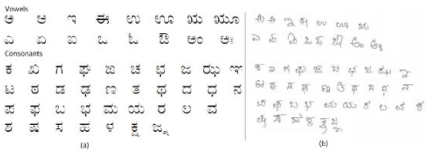


Fig. 1. (Kannada Character set (a)Printed (b)Handwritten (gray scale)

Since, the standard database for handwritten character images is not available in Kannada script; we created our own database for the proposed method. Dataset of handwritten Kannada characters (vowels and consonants) is created by collecting the handwritten documents from writers belonging to different professions. Sample image of Kannada handwritten vowels and consonants, in gray scale, is shown in Fig. 1(b). The collected data sheets were scanned using a flatbed scanner at a resolution of 300 dpi and stored as grayscale images. The noise, due to erratic hand movements and inaccuracies in digitization, present in the image is removed by applying median filter three times. Image binarization is performed using Otsus method. The noise at isolated locations and spikes around the end of the characters are removed using morphological open and close operations. A minimum bounding box is then fitted to the character and

the character is cropped. To bring uniformity among the handwritten characters, each image (located in bounding box) is size normalized to a size of 64×64 pixels. A total of 24500 binary digital images representing Kannada handwritten vowels are thus obtained. Each image represents a handwritten character (binary 1) that is unconstrained, isolated and clearly discriminated from the background (binary0).

3 Feature Extraction

The proposed method is implemented in two stages. In Kannada character set, certain characters are similar in shape and hence are prone to misclassification. Hence, it is essential to perform recognition of a character in two stages. Firstly, similar shape characters are grouped together as shown in Table1. This resulted in 22 classes as compared to original 49 classes [13]. For an input image, features are extracted using crack codes and fed to SVM classifier to classify the input character as to belong to one of these groups. In the second stage, pixel density of the input character is computed and fed to SVM classifier to label the character as to belong to one of the character in that group. In case, the input character belongs to a non-group character as shown in Table 1, the second stage of classification is not performed. The details of feature extraction method are given below.

Table 1. Group of similar shaped characters and other characters

Group of similar shaped characters	
ಅ ಆ	ರ ಠ ಶ ಕ
ಚ ತ	ಹ ಡ ಪ ದ
ಎ ಏ ಐ ವ ಟ	ಮ ಯ ಱ
ಒ ಓ ಖ ಬ ಬ ಜ	ಗ ಸ
ಢ ಛ ಛ ಥ ಧ ಫ ಭ ಷ	

Non-group characters							
ಇ	ಈ	ಉ	ಊ	ಔ	ಂ	ಃ	ಗ
ಞ	ಝು	ಣ	ಲ	ಳ			

3.1 Feature Extraction in the First Stage

The pre-processed character image is divided into non overlapping 64 zones of equal size. For each zone, compute crack code. Crack codes are efficient in representing the region borders. The crack belongs to the boundary that lies between foreground and background pixel. Encoding this line (a sequence of horizontal and vertical pixel edges) yields the crack code of the digitized object boundary (identified as the bold line in Figure 2). Codes are represented with 4 possible directions as shown in Fig. 2.

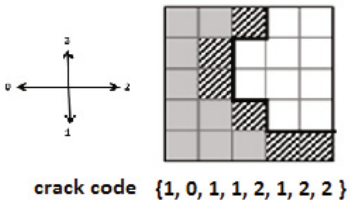


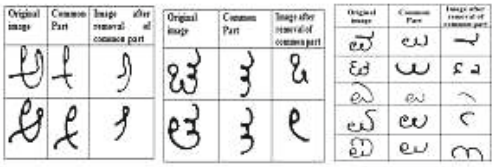
Fig. 2. Crack code (The "crack" is shown with the thick black line)

The crack code of the image is generated by traversing the character in each zone in anticlockwise direction. The method of computing the features using crack codes is presented in our earlier work [18]. The feature set of size 512 is obtained.

3.2 Feature Extraction in the Second Stage

To discriminate between similar shaped characters in the group (Table 1), we employed the concept of object pixel density. As discussed in previous paragraph, characters are grouped based on their shape. In each group, we observe that the common part in the shape does not contribute to the discriminating features. Hence, we decided to eliminate such common parts from the character in that group and compute the features for the remaining part of the character. Examples of such characters are shown in Table 2. Feature extraction is explained below.

Table 2. Character's groups after removal of common part



The common part of the character, for a specific group, is eliminated. Next, traversing the image column-wise from top left to bottom right in each zone, the occurrences of object pixel is counted which gives object pixel density for that zone. Obviously, certain zones have zero value. The feature vector of size 64 is obtained (Fig. 3).

4 Classification

Basically, Support Vector Machines (SVM) [19] is defined for two-class problem and it finds the optimal hyper-plane which maximizes the distance, the margin,

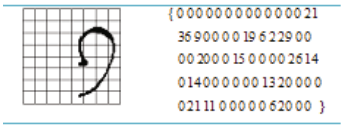


Fig. 3. Zoning of a character and object pixel density

between the nearest examples of both classes, named support vectors (SVs). Given a training database of M data $\{x_m | m = 1, ..., M\}$, the linear SVM classifier is defined as

$$F(x) = \sum \alpha_i x_i x + b \tag{1}$$

where x_j are the set of support vectors and the parameters j and b have been determined by solving a quadratic problem. The linear SVM can be extended to a non-linear classifier by replacing the inner product between the input vector x and the SVs xj, through a kernel function K defined as

$$K(x, y) = \phi(x)\phi(y) \tag{2}$$

This kernel function should satisfy the Mercer's Condition. The performance of SVM depends on the kernel used. A number of simple kernels include polynomial SVM, Radial Basis Function (RBF) SVM, and Two-layer neural network SVM. We have chosen RBF (Gaussian) kernel due to its better performance compared to other kernels. A SVM based classifier contracts all the information contained in the training set relevant for classification, into the support vectors. This procedure reduces the size of the training set by identifying its most important points. The SVM is applied to our multiclass character recognition problem by using one-versus-rest type method. More detailed treatment on principles and applications of SVM can be found elsewhere [20].

5 Experimental Results

Experiments are carried out on 24500 samples of handwritten Kannada characters. The database is divided in to train and test sets. Five-fold cross validation technique is used to test the performance of the proposed method. For each fold 19600 images are used for training and rests of the 4900 images are used for testing. Initially, character set consisting of entire 49 characters is considered and are labeled as 49 classes. Crack codes are used for feature extraction as explained in section 3 and the extracted features of the input character image are fed to trained SVM classifier for labeling the input character image. The average recognition accuracy of 87.24 % is obtained. The misclassification is mainly due to characters that are similar in shape. Taking into account of similar shaped characters, we then performed experiments in two stages as explained in section 3. Totally 9 groups are formed of 36 similar shaped characters and 13 individual classes of non-group characters, thus resulting in 22 classes instead of 49 classes

as shown in Table 1. In the first stage of classification we obtained 94.39% recognition accuracy for 22 classes. Once, the input character was classified to belong to a particular group, the pixel density features of the input character is fed to SVM classifier for character labeling in that group. An example of two stage feature extraction and subsequent classification for the first two vowel characters are shown in tables 3 to 6. It is clear from the tabulated results that, the method of grouping the characters and then implementing the character recognition in two stages provided acceptable recognition rate of 91.02% compared to that obtained without performing groups(Table 6). Further, to the best of our knowledge, results for the entire handwritten Kannada character set for the methods present in the literature are not available for comparison with our proposed method.

Table 3. Confusion matrix of first two vowel characters for 49 classes (not all characters are shown)

	ಕ್ಷ	ಐ	ಜ	ಟ	ಠ	ಢ	ತ	ಪ	ಭ	ರ	ಲ	ಸ	ಳ	ಹ	ಳ	ಅ	ಆ	ಈ	ಋ	ಎ	ಒ	ಓ	ಔ
ಅ	0	0	0	2	2	0	3	0	1	2	3	2	1	1	0	450	27	1	0	1	1	3	0
ಆ	2	1	2	1	2	1	2	1	2	3	0	3	0	0	2	24	450	1	1	0	0	0	2

Taking into account of similar shaped characters, we then performed experiments in two stages as explained in section 3. Totally 9 groups are formed of 36 similar shaped characters and 13 individual classes of non-group characters, thus resulting in 22 classes instead of 49 classes as shown in Table 1. In the first stage of classification we obtained 94.39% recognition accuracy for 22 classes. Once, the input character was classified to belong to a particular group, the pixel density features of the input character is fed to SVM classifier for character labeling in that group. An example of second stage feature extraction and subsequent classification for for the first two vowel characters are shown in tables 3 to 6. It is clear from the tabulated results that, the method of grouping the characters and then implementing the character recognition in two stages provided acceptable recognition rate of 91.02% compared to that obtained without performing groups(Table 6). Further, to the best of our knowledge, results for the entire handwritten Kannada character set for the methods present in the literature are not available for comparison with our proposed method.

Table 4. Confusion matrix of first two vowel characters at (a) stage-1 (obtained from table 4) (b) stage 2

	ಅ	ಆ	ಇ	ಈ	ಉ	ಊ	ಋ	ಎ	ಏ	ಐ	ಒ	ಓ	ಔ	ಕ	ಖ	ಗ	ಘ	ಙ	ಚ	ಛ	ಜ	ಝ	ಞ	ಟ	ಠ	ಡ	ಢ	ಣ	ತ	ಥ	ದ	ಧ	ನ	ಪ	ಫ	ಬ	ಭ	ಮ	ಯ	ರ	ಱ	ಲ	ಳ	ವ	ಶ	ಷ	ಸ	ಹ	ಃ
ಅ	0	6	5	7	0	0	6	1	9	3	2	2	951	0	2	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ಆ																																																

(a) stage-1

	ಅ	ಆ
ಅ	467	10
ಆ	9	465

(b) stage 2

Table 5. Confusion matrix for first two vowel characters (consolidated from tables 4,5)

	ಕ್ಷ	ಜ್ಞ	ಪ್ಪ	ಟ್ಟ	ಠ್ಠ	ಡ್ಡ	ತ್ತ	ಪ್ಪ	ಬ್ಬ	ರ್ರ	ಲ್ಲ	ಸ್ಸ	ನ್ನ	ಹ್ಮ	ಳ್ಳ	ಅ	ಆ	ಈ	ಊ	ಎ	ಒ	ಓ	ಔ
ಅ	0	0	0	2	2	0	3	0	1	2	3	2	1	1	0	467	10	1	0	1	1	3	0
ಆ	2	1	2	1	2	1	2	1	2	3	0	3	0	0	2	9	465	1	1	0	0	0	2

Table 6. Comparison of recognition results of proposed method (consisting of 22 classes in the first stage) with 49 classes in terms of %

Class	49-Classes	2-level Classification
ಅ	90	93.4
ಆ	90	93
ಚ	82.8	98
ಞ	87.4	96
ಡ	81	86.2
ದ	72.6	88.4
ಪ	81.4	88.8
ಹ	82.8	87.6
ರ	87.8	84.4
ಶ	68.6	84.4
ಕ	79.6	88.4
ಳ	89.8	91.6

Class	49-Classes	2-level Classification
ಟ	91.6	98.2
ವ	86.2	93.2
ಏ	92.2	96.2
ಖ	90.4	92.8
ಐ	86.06	94
ಉ	93.8	93.8
ಜ	82	84.6
ಫ	84.8	94
ಬ	87.4	88.2
ಓ	90.4	92.4
ಒ	88.06	90
ಣ	90.2	91.8
ಸ	82.8	96.04

Class	49-Classes	2-level Classification
ಮ	93.04	93.06
ಯ	92.8	96.2
ಱ	93.4	98.4
ಳ	78.2	78.2
ಥ	81.6	81.6
ಧ	77.8	77.8
ಥ	62.8	62.8
ಧ	70.2	70.2
ಘ	68.2	68.2
ಢ	91	91
ಷ	77.6	77.6

6 Conclusion

In this paper we have presented a novel Kannada handwritten character recognition method using SVM classifier. Two types of feature extraction methods are proposed using used crack codes and pixel density. The proposed methodology is implemented in two stages and we have obtained encouraging results. The aim of the proposed system is to remove the confusions among similar shaped classes and thereby increase the recognition rate. This has been demonstrated by performing the experiments on the data set with and without grouping of characters. The proposed method can be extended for characters written in other scripts.

Acknowledgement. The authors are grateful to the referees for their helpful comments. The authors also thank Dr. P. S. Hiremath, Dr. B. V. Dhandra, GUG, and Dr. P. Nagbhushan, UOM, Mysore, for their guidance and support.

References

1. Nagy, G.: Chinese character recognition, A twenty five years retrospective. In: Proceedings of ICPR, pp. 109–114 (1988)
2. Lorigo, L.M., Govindaraju, V.: Offline Arabic handwriting recognition: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(5), 712–724 (2006)
3. Pal, A., Singh, D.: Handwritten English Character Recognition Using Neural Network. International Journal of Computer Science & Communication 1, 141–144 (2010)

4. Indira, K., Sethu Selvi, S.: Kannada Character Recognition System: A Review. Inter JRI Science and Technology 1(2), 31–42 (2009)
5. Pal, U., Chaudhuri.: Indian script character recognition: a survey. Pattern Recognition 37(9), 1887–1899 (2004)
6. Holambe, A.N., Thool, R.C., Jagade, S.M.: Printed and Handwritten Character & Number Recognition of Devanagari Script using Gradient Features. International Journal of Computer Applications 2(9), 975–8887 (2010)
7. Pal, U., Wakabayashi, T., Kimura, F.: Handwritten Bangla Compound Character Recognition Using Gradient Feature. In: 10th International Conference on Information Technology (ICIT 2007), December 17-20, pp. 208–213 (2007)
8. Bhattacharya, U., Shridhar, M., Parui, S.K.: On Recognition of Handwritten Bangla Characters. In: Kalra, P.K., Peleg, S. (eds.) ICVGIP 2006. LNCS, vol. 4338, pp. 817–828. Springer, Heidelberg (2006)
9. Hanmandlu, M., Ramana Murthy, O.V., Madasu, V.K.: Fuzzy Model based recognition of handwritten Hindi characters. In: 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, December 3-5, pp. 454–461. IEEE (2007)
10. Pal, U., Sharma, N., Wakabayashi, T., Kimura, F.: Handwritten Character Recognition of Popular South Indian Scripts. In: Doermann, D., Jaeger, S. (eds.) SACH 2006. LNCS, vol. 4768, pp. 251–264. Springer, Heidelberg (2008)
11. Niranjan, S.K., Kumar, V., Hemantha Kumar, G., Aradhya, M.: FLD based Unconstrained Handwritten Kannada Character Recognition. International Journal of Database Theory and Application 2(3) (September 2009)
12. Dhandra, B.V., Hangarge, M., Mukarambi, G.: Spatial Features for Handwritten Kannada and English Character Recognition. IJCA, Special Issue on RTIPPR (3), 146–151 (2010)
13. Ragha, L.R., Sasikumar, M.: Feature Analysis for Handwritten Kannada Kagunita Recognition. International Journal of Computer Theory and Engineering, IACSIT 3(1), 1793–8201 (2011)
14. Aradhya, M., Niranjan, S.K., Hemant Kumar, G.: Probablistic Neural Network based Approach for Handwritten Character Recognition. Special Issue of IJCCT 1(2,3,4); 2010 for International Conference (ACCTA 2010), August 3-5, pp. 9–13 (2010)
15. Kumar, S., Kumar, A., Kalyan, S.: Kannada Character Recognition System using Neural Network. National Journal on Internet Computing 1, 33–35
16. Gonzalez, R.C.G., Woods, R.E.: Digital Image Processing, 2nd edn. Pearson Education Asia (2002)
17. Alaei, A., Nagabhushan, P., Pal, U.: Fine classification of Handwritten Persian/Arabic Numerals by Removing Confusion amongst Similar Classes. In: 10th International Conference on Document Analysis and Recognition, ICDAR, pp. 601–605
18. Rajput, G.G., Horakeri, R.: Handwritten Kannada Vowel Character Recognition System using Crack code and Fourier Descriptors. In: Sombattheera, C., Agarwal, A., Udgata, S.K., Lavangnananda, K. (eds.) MIWAI 2011. LNCS, vol. 7080, pp. 169–180. Springer, Heidelberg (2011)
19. Vapnik, V.N.: The Statistical Learning Theory. Springer, Berlin (1998)
20. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, pp. 144–152 (1992)

Semi-supervised Clustering by Selecting Informative Constraints

Vidyadhar Rao and C.V. Jawahar

IIIT-Hyderabad, India
{vidyadhar.rao@research.,jawahar@}iiit.ac.in

Abstract. Traditional clustering algorithms use a predefined metric and no supervision in identifying the partition. Existing semi-supervised clustering approaches either learn a metric from randomly chosen constraints or actively select informative constraints using a generic distance measure like Euclidean norm. We tackle the problem of identifying constraints that are *informative* to learn appropriate metric for semi-supervised clustering. We propose an approach to simultaneously find out appropriate constraints and learn a metric to boost the clustering performance. We evaluate clustering quality of our approach using the learned metric on the MNIST handwritten digits, Caltech-256 and MSRC2 object image datasets. Our results on these datasets have significant improvements over the baseline methods like MPCK-MEANS.

Keywords: Semi-supervised Clustering, Constraint Selection, Metric Learning.

1 Introduction

Clustering aims to partition a collection of data items/objects into clusters, such that objects within a cluster are more "similar" to each other than they are to objects in other clusters. A common approach is to infer the similarity based on a distance measure on the data objects. A good distance metric between objects accurately reflects the true underlying relationships i.e., reports small distances for similar objects and large distances for unrelated objects. In literature, the notion of similarity has many variants and formulations [9]; some according to the purpose of the study [8], some making domain-specific assumptions [8] while some relying on prior knowledge of the problem [1] [3] [15].

Semi-supervised clustering algorithms offer a way to specify the prior information in the form of instance level pairwise constraints like must-link constraints and cannot-link constraints. Many works have demonstrated that the use of such constraints could help in guiding the clustering algorithm towards a more appropriate data partitioning. Some of these approaches incorporate the constraints by modifying the clustering objective function so that it includes enforcing constraints [15], or initializing and constraining clustering based on labeled examples [1]. However, it has been shown that constrained clustering is a hard problem [4] and it is not necessarily a good idea to derive the partitions strictly satisfying every constraint [16]. Instead of enforcing the constraints directly, recent techniques introduced penalties on constraint violations.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 213–221, 2013.

Other approaches employ distance metric into the clustering algorithm; where the metric is first trained to satisfy the constraints. The unified approach, MPCK-MEANS [3] uses pairwise constraints along with unlabeled data for constraining the clustering and learning distance metrics. This technique penalizes the constraint violations on must-link and cannot link constraints. However, the constraint specification effort can become a significant burden as they do not offer any mechanism for selecting informative pairwise constraints. Moreover, the cost associated with constraint violations do not include any prior knowledge from the data.

One of the solution to this problem is to adapt an active learning strategy where the algorithm queries an oracle that can assign a must-link or cannot-link label to a given constraint pair. The goal is to query potentially interesting constraint pairs and obtain a better partition of data with minimal number of queries. However, the existing approaches assume that the target metric is available for clustering data. Active PCKMeans [2], depends on farthest-first strategy, can pose a constant number of queries to the oracle by actively selecting informative constraints to get improved clustering performance. Active query selection algorithm [11] is a special case of min-max approach, using a Gaussian kernel to measure the uncertainty in deciding the cluster memberships.

To this end, the previous semi-supervised clustering approaches either learn a metric from randomly chosen constraints or actively select informative constraints using generic distance measures like Euclidean norm or a Gaussian kernel. In this work, we focus on how to utilize the user specified constraints effectively to infer the cluster labels by learning appropriate distance metrics. Unlike MPCK-means, we impose pairwise distance constraints to simultaneously learn metric and perform clustering to guide the process of active constraint set acquisition. We empirically evaluate the effectiveness of our approach on Caltech-256, MSRC2 and MNIST image datasets. The proposed approach performs better than previously proposed methods that either use a pre-defined metric or use random set of constraints for learning a metric.

2 Semi-supervised Clustering

Our goal of learning distance metrics is to improve the clustering performance by selecting informative pairwise constraints. Metric learning often learns the appropriate distance function (metric) from a set of examples in a supervised setting. In a semi-supervised clustering setting, one could learn appropriate metric, if enough pairwise constraints (must-link and cannot-link) are provided. However, getting appropriate constraints from an oracle is often hard. One needs to actively minimize the extra supervision for solving the clustering problem.

Thus, we need to solve these two sub-problems in a coupled manner. Specifically, when the scheme has an access to a set of constraining pairs, our goal is to find potential supervision information and learn appropriate distance metric for semi-supervised clustering.

2.1 K-Means with Mahalanobis Metric

Given N points $X = \{x_i\}_{i=1}^N$, $x_i \in R^d$, we consider the K-means problem as a disjoint K-partition of $X = \{X_h\}_{h=1}^K$. We wish to find the K cluster centroids $\{\mu_h\}_{h=1}^K$ and $l = \{l_i\}_{i=1}^N$, the cluster labels , where $l_i \in \{1, \ldots, K\}$. Points in the same partition are assigned with the same cluster label. The goal is to minimize the objective function

$$\phi_A(X) = \sum_{x_i \in X} d_A(x_i, \mu_{l_i}) \tag{1}$$

where the squared Mahalanobis distance between the points x and y is defined as $d_A(x, y) = (x - y)^T A(x - y)$ and A is a $d \times d$ positive semi-definite matrix. Note that, when $A = I$, the above equation is equivalent to the K-means clustering problem with Euclidean distance metric. Our objective is to learn the distance metric, A using the constraints given by the user and later adopt it in Eqn.(1) so that $\phi_A(X) < \phi_I(X)$.

2.2 Learning with ITML

Information theoretic metric learning (ITML) method of [6] learns a Mahalanobis distance metric, A, under a given set of constraints. Given an initial $d \times d$ positive definite matrix, A_0, specifying the prior knowledge about the inter-point distances, ITML [6] minimizes the LogDet divergence between matrices A and A_0 subject to a set of constraints specified by the user. The goal is to enforce simple distance constraints for similar(S) and dissimilar(D) points and solve the following optimization problem:

$$\begin{aligned} \min\ & D_{ld}(A, A_0) \\ \text{s.t.}\ & A \succeq 0 \\ & d_A(x_i, x_j) \leq u \qquad (i, j) \in S \\ & d_A(x_i, x_j) \geq v \qquad (i, j) \in D \end{aligned} \tag{2}$$

where $D_{ld}(A, A_0) = tr(AA_0^{-1}) - \log det(AA_0^{-1}) - d$; v and u are large and small values, respectively. Solving Eq.(2) involves repeatedly projecting the current solution onto a single constraint, via an update:

$$A_{t+1} = A_t + \beta_t A_t (x_{i_t} - x_{j_t})(x_{i_t} - x_{j_t})^T A_t, \tag{3}$$

where x_{i_t} and x_{j_t} are the constrained data points for iteration t, and β_t is a projection parameter computed by the ITML algorithm. Our goal is to identify informative constraints from the data that minimizes the Eq.(2) to learn the appropriate distance metrics for semi-supervised clustering problem.

2.3 Semi Supervised Clustering by Selecting Informative Constraints

We perform K-means clustering and metric learning using ITML [6] such that both the objectives in Eq.(1) and Eq.(2) are solved simultaneously. We also

provide a quantitative way to measure the constraint-set utility for paritional clustering as adapted in [5].

Informativeness refers to the amount of information in the constraint set that the algorithm cannot determine on its own. *Coherence*, measures the amount of agreement with in the constraints themselves, with respect to a given distance metric. Constraint sets with high informativeness and coherence tend to result in increase in clustering performance [5]. In our approach, we use pairwise distance constraints which are more *informative* than simple must-link and cannot-link constraints. We perform ITML to learn the metric and subsequently use it for the K-means clustering. The main steps for selecting constraints with high *coherence* is summarized in Algorithm 1.

Algorithm 1: Semi Supervised Clustering

Data: X, K, S, D, u, v
Result: A, S_a
$A_0 = I$ *// Initial prior about the inter-point distances*;
$S_u = S \cup D$ *// User specified similar and dissimilar constraints*;
$S_a = \{\}$ *// Active(informative) constraints selected by our algorithm*;
repeat
  foreach $(i, j) \in S_u$ **do**
    //learn metric (A^{ij}) *if a new constraint* $(i, j) \in S_u$ *is added to* S_a
    $A^{ij} \leftarrow \text{ITML}(X, A_0, S_a \cup (i, j), u, v)$;
    // Obtain cluster assignments (l^{ij}) *and quality of clustering* (Q^{ij})
    $(Q^{ij}, l^{ij}) \leftarrow K\text{-means}(X, A^{ij}, K)$; *//use metric* A^{ij} *for* K*-means*
  end
  $(i^*, j^*) \leftarrow \arg\max_{ij}(Q^{ij})$;
  $A_0 \leftarrow A^{i^* j^*}$;
  $S_a \leftarrow S_a \cup (i^*, j^*)$;
  $S_u \leftarrow S_u \setminus (i^*, j^*)$;
until *convergence*;

In each iteration, a new constraint from a set of user specified constraints S_u is added into the active constraint set, S_a. ITML learns the distance metric by using constraints from the active constraint set, S_a. Thereafter, we perform the K-means clustering again using the learned distance metric, A^{ij}, to obtain the cluster labels, l^{ij} and the clustering quality, Q^{ij}. Then update the active constraint set, S_a, with a constraint which resulted in maximum performance in the clustering. We also update S_u accordingly and repeat this process until a satisfactory clustering performance is achieved.

In our experiments, we have selected 50 informative constraints to show the effectiveness of Algorithm 1. In addition to the active constraints generated from Algorithm 1, we can also infer additional constraints using the transitive closure of the set of constraints [2]. Given three data points x, y, z, if $(x, y) \in S$ and $(y, z) \in S$ then $(x, z) \in S$; if $(x, y) \in S$ and $(y, z) \in D$ then $(x, z) \in D$.

Our algorithm requires $O(mn)$ metric learning and clustering operations, where $m = |S_a|$ is the number of informative constraints selected and $n = |S_u|$ is the number of user specified constraints. The complexity of our algorithm scales linearly with the number of user specified constraints. Empirical results suggest only a few informative constraints have significant effect on clustering performance.

3 Experiments and Results

3.1 Experimental Setup

Datasets: To assess the viability of proposed approach, we have performed experiments using MNIST [10] database of handwritten digits, Caltech-256 [7] and MSRC2 [13] object datasets in our experiments. We have selected 11 objects from Caltech-256 dataset and all 20 objects from the MSRC2 dataset for object clustering. For MNIST dataset, we have used all the ten digits for clustering. The parameter K in Algorithm 1, which is the number of clusters in each dataset, is set as the real number of classes in each dataset. Fig.1 shows sample images.

For ITML algorithm, distances are constrained to be either similar or dissimilar, based on the class values, and are drawn only from training set. For MNIST dataset, we choose constraints only from 10 training examples per digit and evaluate the clustering on the 100 testing examples per digit. For the Caltech-256 dataset, we have used 5 images for training and 45 images for testing from each of the 11 classes. For MSRC2 dataset, we have used 5 images for training and 25 images for testing from each of the 20 classes.

Fig. 1. Sample object and handwritten digit images used in our experiments: Caltech-256 (left), MSRC2 (middle), and MNIST (right)

Features: Clustering is performed based on visual features extracted automatically from the images. For the Caltech-256 and MSRC2 object clustering, the image data is supplied in the form of visual words with a vocabulary size of 600 using the popular SIFT descriptors as in VLFeat library [14]. In the MNIST dataset, the digits have been normalized to fit in a 20x20 pixel box and the resulting 400 pixel values are used as feature representation for each image. In our implementation, we empirically set the slack variables v and u in Eqn.(2) to the 95^{th} and 5^{th} percentiles of the distribution of pairwise Euclidean distances within the dataset, respectively.

Evaluation Methods: A number of ways have been developed to validate unsupervised clustering algorithms [17]. In our case, the ground truths for the datasets are naturally available, that is, the digit labels and object categories. We also report the clustering performance by comparing total squared Mahalanobis distances computed from Eq.(1). We also evaluate the quality of clusters using the F_1-measure [2] and Rand Index [12] commonly used performance metric for semi-supervised clustering algorithms.

3.2 Results and Discussion

We compare our approach to the baseline methods like popular K-means which is an unsupervised clustering with Euclidean distance metric and MPCK-MEANS which is a semi-supervised clustering algorithm that simultaneously learns metric under some constraints. We quantify the comparison by using the same initial centroids for K-means in all the methods. For semi-supervised algorithms, we show the results by selecting 50 informative constraints using Algorithm 1.

Three variants of Algorithm 1 are implemented: Semi-supervised clustering with random constraints (SSC-rand), Semi-supervised clustering with online distance metric learning (SSC-OLDML) and Semi-supervised clustering with active constraint set generation (SSC-active). In SSC-rand, we perform K-means clustering with a distance metric learned directly from random constraints. In SSC-OLDML, we perform K-means clustering with a distance metric learned using the most recently obtained metric as prior (using Eq.(3) for ITML). In SSC-active, we perform K-means clustering with a distance metric learned using the active constraint set acquired in each iteration (using Eq.(2) for ITML).

We present experimental results of our approach on three image datasets and compare them with the baseline methods. Table 1 shows the K-means error for all methods. The results show that the SSC-active performs better than popular K-means (does not learn a metric), SSC-rand (learns metric from random constraints) and SSC-OLDML methods. We have not included the MPCK-means algorithm in Table 1, as it includes penalties for constraint violations in the objective function.

Table 1. Unsupervised K-means clustering error, $\phi_I(.)$, along with Semi-Supervised clustering error, $\phi_A(.)$, with 50 informative constraints

	Algorithm			
Dataset	Popular K-means	SSC-rand	SSC-OLDML	SSC-active
MNIST	37380	36562	61474	**34726**
Caltech-256	2.665	2.565	2.618	**2.020**
MSRC2	2.059	2.275	3.344	**1.991**

In Table 2, we have used rand index [12] to evaluate the semi-supervised clustering algorithms. We notice that MPCK-means does not boost the performance when compared to popular K-means on MNIST and MSRC2 datasets. This shows that incorporating random constraints might degrade the clustering

performance and it is very critical to choose constraints that are informative for semi-supervised clustering.

SSC-active adopts distance metric learning by choosing informative constraints, and always performs better than the unsupervised K-means on all three datasets. The underlying reasoning seems to be that the pairwise distance constraints are more *informative* and also that the constraints selected by our approach have high *coherence* with respect to the learned metric.

In contrast to what we observed for MNIST and MSRC2 datasets, SSC-active did not quite perform well on the Caltech-256 images in comparison to the MPCK-means. The images in Caltech dataset typically share multiple object categories and we believe that this can be surmounted by use of large vocabularies or using more features for image representation.

Table 2. Results of the semi-supervised clustering based methods, measured in Rand Index (higher is better)

	Algorithm				
Dataset	Popular K-means	SSC-rand	SSC-OLDML	MPCK-means	SSC-active
MNIST	0.875	0.881	0.861	0.862	**0.921**
Caltech-256	0.769	0.758	0.827	**0.841**	0.807
MSRC2	0.892	0.895	0.881	0.859	**0.904**

In Table 3, we compare the performance of clustering methods in terms of F_1-measure [2]. The results demonstrate that our approach performs close to the MPCK-MEANS on the object datasets and outperforms on handwritten digits images. However, we can see that SSC-active always performs better than the unsupervised clustering on all three datasets.

Table 3. Comparison of the different semi-supervised methods on three datasets, measured in F_1 score (higher is better). For each dataset, we see that SSC-active performs better than unsupervised clustering and performs close to MPCK-MEANS.

	Algorithm				
Dataset	Popular K-means	SSC-rand	SSC-OLDML	MPCK-means	SSC-active
MNIST	0.410	0.434	0.334	0.377	**0.621**
Caltech-256	0.150	0.156	0.195	**0.249**	0.215
MSRC2	0.155	0.162	0.128	**0.226**	0.203

The *SSC-OLDML* approach sometimes degrades the clustering performance (See Tables 1, 2 and 3). This happens because of the greedy fashion in which the metric has been learned; using most recently learned metric as prior can constrain later ones due to potential conflicts between the prior metric and the new constraints, and there is no mechanism for backtracking. From this we can infer that, the constraint set generated by *SSC-OLDML* has low *coherence* and therefore can lead the algorithm into unpromising areas of search space. However, the *SSC-active* continued to be consistent as it learns the metric from all the constraints using the prior as $A_0 = I$ and thereby acquiring the constraints sets

with high *coherence* i.e., amount of agreement within the constraints themselves is high with respect to the learned distance metric.

Overall, our results show that the coupled approach (SSC-active), to utilize informative constraints and learn metric, can boost the performance of semi-supervised clustering. Our approach required only a very small number of informative constraints to gain significant improvements in the clustering over the existing semi-supervised clustering approaches like MPCK-means.

4 Conclusion

Pairwise constraints would facilitate accurate metric learning and boost the quality of semi-supervised clustering algorithms. This paper has presented an approach to jointly learn metric and select informative constraints from a given set of pairwise distance constraints. We partition the entire data using K-means with the learned metric. Our semi-supervised algorithm was applied on the image datasets and its application always achieved better than the baseline methods like unsupervised K-means with Euclidean metric and semi-supervised MPCK-means clustering. Our results demonstrate that the more informative constraints are under the learned metric, the more likely they are to improve clustering.

References

1. Basu, S., Banerjee, A., Mooney, R.: Semi-supervised clustering by seeding. In: Machine Learning-International Workshop then Conference, pp. 19–26 (2002)
2. Basu, S., Banerjee, A., Mooney, R.J.: Active semi-supervision for pairwise constrained clustering. In: SDM, pp. 333–344 (2004)
3. Bilenko, M., Basu, S., Mooney, R.J.: Integrating constraints and metric learning in semi-supervised clustering. In: ICML, p. 11. ACM (2004)
4. Davidson, I., Ravi, T.: Clustering with constraints: feasibility issues and the fk-means algorithm. In: SDM, pp. 138–149 (2005)
5. Davidson, I., Wagstaff, K.L., Basu, S.: Measuring constraint-set utility for partitional clustering algorithms. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 115–126. Springer, Heidelberg (2006)
6. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML, pp. 209–216. ACM (2007)
7. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (2007)
8. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognition Letters 31(8), 651–666 (2010)
9. Jain, A.K., Dubes, R.C.: Algorithms for clustering data. Prentice-Hall, Inc. (1988)
10. LeCun, Y.: Mnist dataset (2000), `http://yann.lecun.com/exdb/mnist/`
11. Mallapragada, P.K., Jin, R., Jain, A.K.: Active query selection for semi-supervised clustering. In: ICPR, pp. 1–4. IEEE (2008)
12. Rand, W.M.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical association 66(336), 846–850 (1971)
13. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)

14. Vedaldi, A., Fulkerson, B.: Vlfeat: An open and portable library of computer vision algorithms. ACM Multimedia, 1469–1472 (2010)
15. Wagstaff, K., Cardie, C., Rogers, S., Schrödl, S.: Constrained k-means clustering with background knowledge. In: Machine Learning-International Workshop then Conference, pp. 577–584 (2001)
16. Wagstaff, K.L., Basu, S., Davidson, I.: When is constrained clustering beneficial, and why? Ionosphere 58(60.1), 62–63 (2006)
17. Xu, R., Wunsch, D.: Clustering, vol. 10. Wiley-IEEE Press (2008)

Spatio-temporal Change Detection of Urban Heat Islands Using Spatial Interpolation

Shrutilipi Bhattacharjee[1], Titas Aikat[2], and Soumya K. Ghosh[3]

School of Information Technology, Indian Institute of Technology Kharagpur, India
{shrutilipi.2007,titasaikat}@gmail.com, skg@iitkgp.ac.in

Abstract. The urban heat island has drawn significant attention in recent years. It has grabbed an unprecedented research interest in the field of remote sensing and geographic information system. This work focuses on spatial interpolation based change detection of heat islands and finding the dynamics of the heat islands. Performance checking is done on derived spatial attribute, land surface temperature (LST). The LSTs of two distinct Landsat ETM+ satellite images (of years 2000 and 2010) are used to verify the accuracy of the proposed algorithm.

Keywords: Urban Heat Island, Geographic Information System, Spatial Interpolation, Inverse Distance Weighting.

1 Introduction

The 20th century is witnessing "the rapid urbanization of the world's population", as the global proportion of urban population is rising dramatically [1]. It is well known and documented that urbanization can have significant effects on local weather and climate. Among these effects, one of the most important is the urban heat island (UHI), which is the direct representation of environmental degradation. An urban heat island is a geographic region which is significantly warmer than its surrounding areas. The higher urbanization leads to more distinct urban heat island with significant temperature differences between urban and rural areas.

Satellite remote sensing is widely used to detect the growth of heat islands, their spatio-temporal drifting, update existing maps, etc. In recent years, many change detection tools and techniques, such as, *visual interpretation*, *pixel-based methods* and *object-based approaches* etc. have been developed. The remote sensing satellite images are in the form of raster data, whereas, geographic information system (GIS) mostly deals with vector data in form of *geographic markup language* (GML) *file*, *shapefile*, etc. Further, GIS layers are formed by derived spatial attributes, such as, *land surface temperature* (LST), *moisture stress index* (MSI) etc. So, image analysis is not always a good choice for GIS. The proposed scheme of UHI change detection primarily focuses on the spatio-temporal drifting of the point/or the spatial region responsible for drastic change of temperature, refered as UHI, and finding the dynamics of these UHI points and their temporal path of drifting. Two derived attribute datasets can be analysed to capture

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 222–228, 2013.

the change from one dataset to another. The reference dataset is considered with respect to which the change detection has been carried out in the second (target) dataset. It is done through temporal mapping of the spatial interpolation technique from the reference dataset to the target dataset. In this work, LSTs, derived from Landsat ETM+ image datasets (source: USGS[1]) are used for analysis and one of the most popular interpolation techniques, *inverse distance weighting (IDW)* [2], is considered for detecting drift of UHI points.

UHIs can be divided into two types, based on the methods of temperature measurements, namely, the *canopy layer heat island* and the *boundary layer heat island* [3]. Several literature surveys have been reported regarding detailed description of UHI formation and detection, their effect analysis, strategies to reduce urban heat island etc. [4] [5]. Oleson *et al.* [6] analysed UHI characteristics in details with respect to a global climate model. Weng *et al.* [7] have studied the urban heat island (UHI) phenomenon to examine the spatial distribution of surface radiant temperatures in Guangzhou, China. Yuan *et al.* [8] investigated the relationships between the *land surface temperature* (LST), percent impervious surface area (ISA), and the normalized difference vegetation index (NDVI) to compare the ISA and NDVI as the influential parameter for surface UHI effects. Characteristics of UHI and boundary layer structures were analysed in [9] using conventional and Moderate Resolution Imaging Spectroradiometer (MODIS) observations, taking Beijing area, China as the example scenario case study.

The propsed work analyses the derived attribute information from satellite images for detecting spatio-temporal drifting of UHI points and finding their track of movement. The main objectives of this work are as follows,

- Study of LST for Landsat ETM+ image datasets.
- Detection of spatio-temporal drifting of UHI points from reference dataset to the target dataset.
- Analysing the temporal movement of UHI points and finding the drifting path.

The paper is organized as follows. Section 2 gives detailed description of the proposed framework. Case study with the proposed algorithm with ETM+ image datasets, is specified in the Section 3. Finally the conclusion is drawn in the Section 4.

2 Framework to Analyse UHI Drifting

The proposed framework deals with the derived spatial attribute *land surface temperature*, LST, which is present in the geographic data repository in the form of vector data. Finding spatio-temporal drifting point of the heat islands in the target dataset, is based on spatial interpolation technique with respect to the reference dataset.

For this work, it is assumed that the UHI points in the reference dataset are known. LST values of those UHI points are plotted and a curve of LSTs is

[1] `http://www.usgs.gov/`

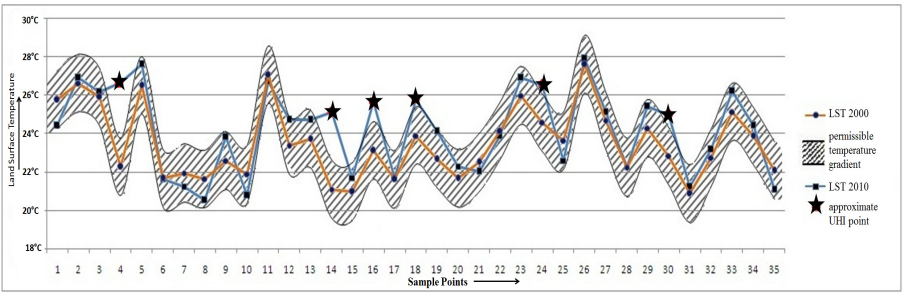


Fig. 1. LST at each UHI for the reference and the target datasets with permissible temperature gradient as buffer

obtained (refer Fig. 1). Taking the average temporal change of LST value from the reference year dataset to the target year dataset, a buffer is drawn along the curve. It is denoted as the permissible temperature gradient in Fig. 1. The target dataset where the drift of the newly developed UHI points is to be predicted, are identified which represents the same coordinate points as the reference dataset. They are considered as the candidate UHI points in the target dataset. LSTs of the target dataset at those points are also plotted accordingly. If the curve corresponding to any of the UHI point is changed more than threshold (i.e. above the permissible temperature gradient), this point is taken as the newly found candidate UHI point. It is shown in Fig. 1 where, UHI point ids: 4, 14, 16, 18, 24 and 30 (given in the X axis, identified with the symbol $\star$) are selected as the candidate/ approximate UHI points in the target dataset.

The UHIs which are selected as the candidate UHI points are analysed further for drift detection. The overall framework is depicted in Fig. 2. To find the drifted UHI points in the target dataset, the selected i^{th} UHI point is then represented by its surrounding points as per spatial interpolation technique, *IDW*. A predefined number of interpolating points are chosen against each candidate UHI point. The LST of selected UHI point is represented as the weighted sum of its surrounding interpolating points. So, i^{th} UHI point in both reference and target dataset can be represented as,

$S_{1i} = w_{1i1}.K_{1i1} + w_{1i2}.K_{1i2} + ... + w_{1ij}.K_{1ij}$

$S_{2i} = w_{2i1}.K_{2i1} + w_{2i2}.K_{2i2} + ... + w_{2ij}.K_{2ij}$

where S_{1i} and S_{2i} represent i^{th} UHI point's LST of the reference and the target dataset, respectively. The K_{mij} and the w_{mij} represent LST of the j^{th} interpolating point and its weight assigned by the *IDW* with respect to i^{th} UHI point, repectively. Here, $m = 1$ and 2 for the reference and the target datasets, respectively. Now, as per *IDW*, w_{1ij} and w_{2ij} are the functions of distance and are same for any j. Difference in S_{1i} and S_{2i} causes due to change in LST values of its surrounding interpolating points, i.e. for the difference of K_{1ij} and K_{2ij}, for all j.

Now, each pair of K_{1ij} and K_{2ij} for i^{th} UHI point are examined to find the maximum difference above a predefined threshold. The j^{th} interpolating point at which maximum change is found, is taken into consideration and analysed

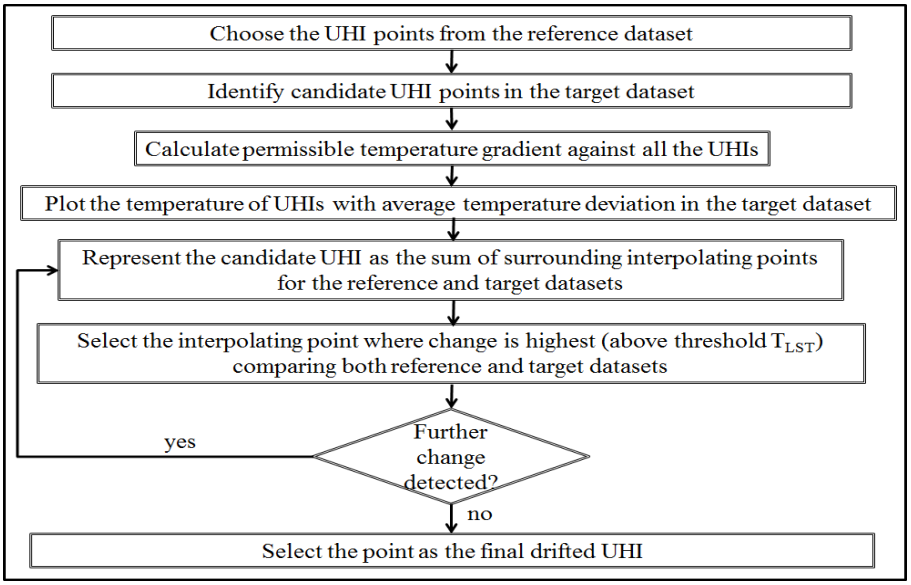


Fig. 2. Process flow to select drifted UHI point in the target dataset

further. This process iterates until no change found or after certain number of iterations are done corresponding to a particular UHI point. Finally, the selected coordinate point where the iteration stops is selected as the drifted UHI point with respect to the reference dataset. This process will also help in finding the temporal track of movement of UHI points. The selected UHI point in each of the iterations will be further connected to get the probable path of movement for the UHI points. The same process can be applied for all the candidate UHI points. As the number of points are finite, this process is also finite. The mechanism is shown in Algorithm 1.

3 Case Study

For the performance study and result analysis, an experiment has been carried out with derived LST data of the year 2000 (reference dataset) and 2010 (target dataset) of *Kolkata* which is one of the major metropolitan cities in India (central coordinate: 22.567°N 88.367°E). Six UHI points are taken from the reference dataset. The drifted UHI points are identified using the Algorithm 1. For example, in Fig. 3, point no. 30 is taken for further analysis using Algorithm 1. At initialization phase, it is labelled as point 30. After first iteration, newly selected candidate UHI is point no. 30_1, then point no. 30_2 and so on. After 3^{rd} iteration, no further cange is found. Point no. 30_3 is selected as the final drifted UHI point corresponding to the candidate UHI no. 30. The probable path of movement of the UHI point no. 30 can be given as, $30 \rightarrow 30_1 \rightarrow 30_2 \rightarrow 30_3$ (refer Fig. 3). Though it is shown in distinct Landsat ETM+ satellite image,

Algorithm. Find drifted UHI point from candidate UHI point
Input: Candidate UHI point S_{2i}, maximum number of iterations: n, number of interpolation points against each candidate UHI point: m
Output: Drifted UHI point A_i

```
A_i = S_2i;
while number of iterations for A_i < n do
    max = |S_2i − S_1i|;
    foreach j = 1 → m do
        if |K_2ij − K_1ij| > max then
            max = |K_1ij − K_2ij|;
            p = j;
        end
    end
    if max == |S_2i − S_1i| then
        Select new m interpolating points against i^th UHI point
    end
    ; else
        A_i = p^th interpolating point of i^th UHI point representing S_2i
    end
    ;
end
Output A_i as the final drifted UHI point
```

Algorithm 1. Detection of drifted UHI point from the candidate UHI point

the actual computation is done on derived data. Analysis of six UHI points are shown in Table 1. The candidate and the drifted UHI point's coordinate information are recorded against each of the UHI points.

Discussion: The proposed scheme is efficient compared to pixel-by-pixel analysis. Typically pixel-by-pixel analysis is time consuming and often requirs expert's intervention. For the proposed methodology, number of pixels to be analysed and the number of iterations are less. In case of the given example, Algorithm 1 takes maximum 10 iterations for the candidate UHI point detection. Though raw Landsat image is not required for drifted UHI point detection through proposed scheme, for checking the accuracy of this scheme, two images and their corresponding derived attribute information are taken for the case study.

4 Conclusion

The urban heat islands are becoming a major threat in today's world. It has adverse effect in terms of global warming, air pollution, human health, etc. The proposed framework focuses on spatio-temporal drifting of urban heat islands. It may be useful to find the pattern of movement of UHIs to take necessary actions. This methodology is found to be efficient with a very few number of sample points. The probable path of temporal drifting of UHI points can be detected by the proposed approach. This work can be extended further for investigating the

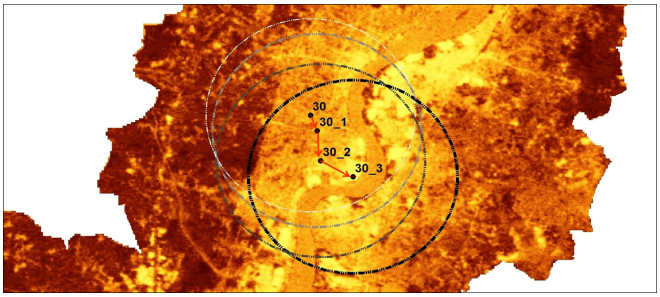


Fig. 3. Drifting of a UHI point from its candidate UHI location

Table 1. Experimentation details using Algorithm 1 on Landsat ETM+ images

Candidate UHI Point ID	Candidate UHI Coordinate Location	Final Drifted UHI Coordinate Location	Temperature Difference (°C)
1	(22.5686°N, 88.1824°E)	(22.5733°N, 88.1772°E)	11.96
2	(22.8064°N, 88.3173°E)	(20.8057°N, 88.3120°E)	11.66
3	(22.9686°N, 88.4591°E)	(22.9585°N, 88.4652°E)	9.59
4	(22.8881°N, 88.3970°E)	(22.8826°N, 88.4128°E)	9.09
5	(22.5939°N, 88.3213°E)	(22.6013°N, 88.3079°E)	7.60
6	(22.8497°N, 88.3586°E)	(22.8323°N, 88.3703°E)	7.75

drifting of UHI points with the growth of urbanizations, abatement of vegetation cover etc.

References

1. United Nations World urbanization prospects: The 2005 revision (2006)
2. Li, J., Heap, A.D.: A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics 6(3), 228–241 (2011)
3. Hekster, F., De Voogt, P., Pijinenburg, A., Laane, R.: Perfluoroalkylated substances: Aquatic environmental assessment. Rijksinstituut voor Kust en Zee/RIKZ 99 (2002)
4. Rizwan, A.M., Dennis, L.Y., Liu, C.: A review on the generation, determination and mitigation of urban heat island. Journal of Environmental Sciences 20(1), 120–128 (2008)
5. US EPA: Draft Report: In reducing urban heat islands: Compendium of strategies. Urban Heat Island basics (2003)
6. Oleson, K.W., Bonan, G.B., Feddema, J., Jackson, T.: An examination of urban heat island characteristics in a global climate model. International Journal of Climatology 31(12), 1848–1865 (2011)

7. Weng, Q.: Fractal analysis of satellite-detected urban heat island effect. Photogrammetric Engineering and Remote Sensing 69(5), 555–566 (2003)
8. Yuan, F., Bauer, M.E.: Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in landsat imagery. Remote Sensing of Environment 106(3), 375–386 (2007)
9. Miao, S., Chen, F., LeMone, M.A., Tewari, M., Li, Q., Wang, Y.: An observational and modeling study of characteristics of urban heat island and boundary layer structures in beijing. Journal of Applied Meteorology and Climatology 48(3), 484–501 (2009)

Distance based Incremental Clustering for Mining Clusters of Arbitrary Shapes

Bidyut Kr. Patra[1,3,⋆], Ville Ollikainen[1], Raimo Launonen[1], Sukumar Nandi[2], and Korra Sathya Babu[3]

[1] VTT Technical Research Centre of Finland, Espoo, FI-02044 VTT, Finland
[2] Indian Institute of Technology Guwahati, Guwahati, Assam, PIN-781039, India
[3] National Institute of Technology Rourkela, Rourkela, Odisha, PIN-769 008, India
{ext-bidyut.patra,ville.ollikainen,raimo.launonen}@vtt.fi,
sukumar@iitg.ernet.in, ksathyababu@nitrkl.ac.in

Abstract. Clustering has been recognized as one of the important tasks in data mining. One important class of clustering is distance based method. To reduce the computational and storage burden of the classical clustering methods, many distance based hybrid clustering methods have been proposed. However, these methods are not suitable for cluster analysis in dynamic environment where underlying data distribution and subsequently clustering structures change over time. In this paper, we propose a distance based incremental clustering method, which can find arbitrary shaped clusters in fast changing dynamic scenarios. Our proposed method is based on recently proposed al-SL method, which can successfully be applied to large static datasets. In the incremental version of the al-SL (termed as *IncrementalSL*), we exploit important characteristics of al-SL method to handle frequent updates of patterns to the given dataset. The IncrementalSL method can produce exactly same clustering results as produced by the al-SL method. To show the effectiveness of the IncrementalSL in dynamically changing database, we experimented with one synthetic and one real world datasets.

Keywords: Incremental clustering, arbitrary shaped clusters, large datasets.

1 Introduction

Clustering has been recognized as one of the important tasks in data mining for data analysis. Clustering activity is to find groups of patterns, called clusters in a dataset $\mathcal{D}$, in such a way that patterns in a cluster are more similar to each other than the patterns in distinct clusters. Clustering methods are mainly divided into two categories based on the way they produce clusters *viz., partitional clustering* and *hierarchical clustering* methods. Partitional clustering methods create single clustering (also known as flat clustering) of a dataset. The partitional methods can be further categorized into two classes based on the criteria used *viz., distance based* and *density based*. Distance based methods optimize a global criteria based on the distance between patterns. Some of the

⋆ This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. The research leading to these results has received funding from the European Union Seventh Framework Programme ($FP7/2007-2013$) under grant agreement 246016.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 229–236, 2013.

popular distance based clustering methods are k-means, CLARA, CLARANS. Density based partitional clustering methods optimize local criteria based on density distribution of patterns. DBSCAN [1] is a popular example of density based partitional clustering method. Hierarchical clustering methods create a sequence of nested clusterings of a dataset. These methods can also be categorized into two classes *viz.,* distance based and density based. The single-link method [2] is a distance based hierarchical clustering method, which can find arbitrary shaped clusters in many applications.

Recent advances in storage, network and computer technology result in producing massive size of databases. To analyze clustering structures in these databases (datasets), several approaches have been proposed [3,4]. One prominent category is hybrid clustering in which a linear partitional clustering is combined with a hierarchical clustering method to reduce the computational complexity of the hierarchical method with compromising little clustering quality [5,6,7,8] or without compromising clustering quality [9,10]. However, this approach is not suitable for dynamic databases as clustering structure change over updations of the dataset. One cannot afford to apply the hybrid clustering repeatedly from scratch whenever there is an update in the database. One way to handle this problem is to adopt these clustering methods in dynamic scenarios such as purchasing patterns of a new product, financial transactions, early detection of epidemic, *etc*.

In this paper, we propose a distance based incremental clustering method which can find clusters of arbitrary shapes and sizes in fast changing databases. Our incremental clustering is based on scalable distance based clustering *al*-SL, which can find arbitrary shaped clusters in metric databases [10]. We restrict our search space for potential changes of clustering membership of patterns after inclusion of new points in the dataset using leaders clustering and metric space properties. Experimental results demonstrate the efficiency of the proposed incremental clustering (termed as IncrementalSL) in dynamic datasets.

The rest of the paper is organized as follows. Section 2 describes a summary of related works. Section 3 discusses the *al*SL clustering method. Section 4 describes the proposed IncrementalSL method. The experimental results and conclusions are discussed in Section 5 and Section 6, respectively.

2 Related Work

Many incremental clustering methods have been developed by many researchers over the decades. The study of incremental clustering can be traced back to late 70's in the last century [11,12]. The leaders clustering is a distance based linear incremental clustering method. For a given threshold distance τ, it produces a set of leaders $\mathcal{L}$ incrementally. For each pattern x, if there is a leader $l \in \mathcal{L}$ such that $||x - l|| <= \tau$, then x is assigned to the cluster represented by l. If there is no such leader, then x becomes a new leader. Each leader can be seen as a representative of a cluster of patterns which are grouped with it. However, it cannot find arbitrary shaped clusters and it is an order depended clustering method. Other variant of leaders method can be found in [13].

It may be noted that clustering in incremental databases has significant different from the clustering applied to data stream. A clustering algorithm in data streams works

in different restricted scenarios. An incremental clustering can have access the whole dataset along with the new updates whereas data stream algorithm can view a limited number of objects at a given time.

The BIRCH clustering proposed in [14] has an incremental phase called *Clustering Feature (CF)*, which utilizes vector space (Euclidean space) properties to store summarized information of a micro cluster of k data points incrementally. Charikar et al. in [15] developed a randomized incremental clustering method, which minimizes maximum diameter of a cluster. Chen et al. [16] introduced an incremental hierarchical clustering method GRIN for numerical datasets, which is based on the gravity theory in physics. In the first phase it uses an agglomerative hierarchical clustering method to produce clustering hierarchy (dendogram) of the given dataset. Then it restructures the hierarchy by merging the lower levels. Each leaf node in the restructured dendogram is represented by the centroid, radius and number of points formed the leaf node. In the second phase, a new data points is either inserted to a leaf node or gravity theory is applied to decide the proper position of the new point. The method may apply the agglomerative hierarchical method to the new set of points if its number exceeds a predefined threshold and follow same restructuring process. However, these methods are not suitable for non Euclidean datasets.

Widyantoro el. al [17] proposed an incremental agglomerative hierarchical clustering, which also uses building and restructuring process of the dendogram of the datasets. It uses minimum spanning tree (MST) to find neighborhood of each data point and it uses vector space property to represent centre of the neighborhood. It is a computationally expensive approach, which cannot be used in large dynamic databases.

IncrementalDBSCAN [18] is a first density based incremental clustering method, which is based on the popular density based clustering DBSCAN [1]. Due to density-based characteristics of DBSCAN, the effects of insertions and deletions are restricted to neighborhood of these data points. Initially, classical DBSCAN is applied to the available dataset at a point of time. Then, IncrementalDBSCAN finds the neighborhood objects and its affected objects in successive updates of the datasets. However, both methods (DBSCAN AND IncrementalDBSCAN) needs two input parameters, which are very difficult to find. The DBSCAN is known an expensive clustering method. IncrementalDBSCAN also needs on an average more than one region quarries for an incremental update.

3 The *al*-SL Clustering Method

Our proposed incremental clustering method IncrementalSL is based on recently proposed distance based *al*-SL method. We choose to adopt this method in dynamic scenarios because the *al*-SL method needs only a single input parameter (h), which is the inter-cluster distance and the value of h can be estimated using approaches given in [10]. The *al*-SL can find arbitrary shaped clusters in large datasets and datasets can be from any metric space (vector or non vector space). In this section, a brief description of the *al*-SL method is reported.

The *al*-SL clustering method is a two-phase clustering method. In the first phase, a hybrid clustering method called *l*-SL, which is the combination of leaders clustering

Algorithm 1. *al*-SL($\mathcal{D}$, h)

Apply leaders clustering method with threshold $\tau = h/2$ to $\mathcal{D}$.
Apply SL method to the leaders set.
Identify a set of pairs of clusters (subsequently leaders) with distance less than $2h$
Apply the leaders clustering method and store followers of above mentioned leaders.
Find two nearest patterns x and y from each pair of clusters with distance less than $2h$.
if $||x - y|| \leq h$ **then**
  Merge the pair of clusters into a single cluster;
end if
Each leader in each cluster is replaced by its followers set. This gives a clustering of the dataset.

method and classical single-link (distance h stopping condition called SL method) is applied to a large dataset. However, the clustering results produced by *l*-SL method may deviate from the final clustering produced by the SL method applied directly to the dataset. A few clusters in the final results of *l*-SL method may be required to be merged in order to obtain the exact final results as that of the SL method. Therefore, the *al*-SL method merges the required clusters in the second phase. The method is reproduced in Algorithm 1. Objects in a cluster produced by *al*-SL method are related to each other by a binary relation 'reachable', which is formally defined as follows.

Definition 1 (Reachable) *A pattern $x \in \mathcal{D}$ is reachable from another pattern $y \in \mathcal{D}$, if there exists a sequence of patterns $x_1, x_2, \ldots, x_p, \quad x_1 = y, x_p = x$ such that $||x_i - x_{i+1}||_{i=1..p-1} <= h, x_i \in \mathcal{D}$.* ■

This 'reachable' relation $R \in \mathcal{D} \times \mathcal{D}$ is an equivalence relation. Therefore, a cluster C in *al*-SL method is an equivalence class of $\mathcal{D}$ by R, *i.e.* $C = [x]_R = \{y \in \mathcal{D} \mid xRy\}$.

4 IncrementalSL: Proposed Incremental Clustering Method

The *al*-SL clustering method works with static large datasets. However, financial transactional database, Recommendation engine database, birth/death registration databases in populated country like India, China are large and dynamic in nature, *i.e.,* these databases have frequent updates of objects. Therefore, clustering structures of these datasets change over time. To capture changes in clustering structures, we present a novel incremental approach called IncrementalSL in this section.

Proposed IncrementalSL updates clustering structures whenever new points are added to the datasets. The method starts with the clustering output produced by the *al*-SL method at a certain point of time. A new object x may influence more than one existing clusters and merge them into one. The set of objects which are influenced over insertion of an object x can be defined as follows.

Definition 2 (Influenced objects) *Let π_{alSL} be the clustering obtained by applying* al-*SL method over the dataset $\mathcal{D}^t$ and x be an object to be inserted to $\mathcal{D}^t$. The set of objects influenced by x is*

$INFLO(x) = \{y \in \mathcal{D}^t \mid xRy,$ *R is a reachable relation on* $\mathcal{D}^t \cup \{x\}\}$. ■

The point x may change the cluster membership of objects in $INFLO(x)$. Finding $INFLO(x)$ is computationally expensive. Therefore, we use leaders and their cluster memberships to find the set $INFLO(x)$. It may be noted that a cluster in *al*-SL consists of a number of leaders with threshold $h/2$, where h is the cluster separation. We only retrieve some key leaders of $\mathcal{D}$ to update the cluster membership information of $INFLO(x)$ instead of entire set $INFLO(x)$. We introduce the following definition to identify the key leaders, which are responsible for possible change in clustering structures.

Definition 3 (Key Leaders) *Let π_{alSL} be the clustering obtained by applying* al-*SL method with parameter h over the dataset $\mathcal{D}^t$ and x be an object to be inserted to $\mathcal{D}^t$. The set of leaders influenced by x called Key Leaders set is*

$Key\_Leaders(x) = \{l \in \mathcal{D}^t \mid ||l - x|| \leq 1.5h, l \text{ is a leader in } \mathcal{D}^t\}$. ■

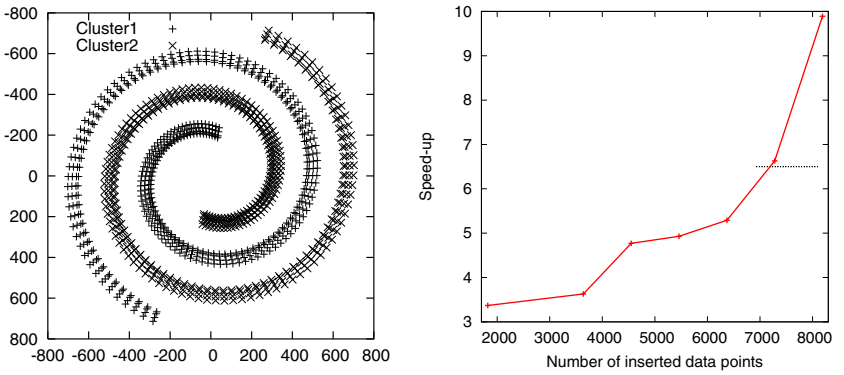
The key leaders of x is the set of leaders located not more than $1.5h$ distance away from x. We use leaders in $Key\_Leaders(x)$ to update the membership information of $INFLO(x)$. We utilize triangle inequality property to quickly locate non-influenced regions in the dynamic datasets using the following Lemma 1.

Lemma 1. *Let l be the closest leader to a newly inserted data point x. If distance between x and l is more than $1.5h$, then no cluster in the dataset $\mathcal{D}$ will be affected due to this insertion.*

Proof: Let y be an arbitrary follower of the leader l. As y is a follower of l, we get $||l - y|| \leq h/2$. By assumption, we get $||l - x|| > 1.5h$. Using triangle inequality, one can write, $||x - y|| + ||l - y|| \geq ||l - x||$. So, $||x - y|| \geq ||l - x|| - ||l - y|| > h$. There is no point in $\mathcal{D}$, which is reachable from x. Therefore, $INFLO(x) = \{x\} \nsubseteq \mathcal{D}$. ■

The IncrementalSL method works as follows while inserting an object x into the dataset $\mathcal{D}$. It calculates distance from x and all leaders in the dataset $\mathcal{D}$. Let $l_{closest}$ be the closest leader with cluster membership C_i.

I. If the distance between x and $l_{closest}$ is less than or equal to $h/2$, the point becomes follower of the leader $l_{closest}$ and x is merged with the cluster in which $l_{closest}$ belongs. Identify the $Key\_Leaders$ of x and if distance between x and a follower of a leader $l \in Key\_Leaders \backslash \{l \in C_i\}$ is less than or equal to h, then C_i is merged with the cluster in which l belongs. Similarly, C_i may be merged with more than one clusters if each of them has a follower f' such that $||x - f'|| \leq h$.
II. If distance between x and $l_{closest}$ satisfies $h/2 < ||x - l_{closest}|| \leq h$, then x becomes a new leader and its membership is C_i. Identify the $Key\_Leaders$ of x and if distance between x and a follower of a leader $l \in Key\_Leaders \setminus \{l \in C_i\}$ is less than or equal to h, then C_i is merged with the cluster in which l belongs. Similarly, C_i may be merged with more than one clusters if each of them has a follower f' such that $||x - f'|| \leq h$.
III. If distance between x and $l_{closest}$ satisfies $h < ||x - l_{closest}|| \leq 1.5h$, then x becomes a new leader. Identify the $Key\_Leaders$ of x and if distance between x and a follower of a leader $l \in Key\_Leaders$ is less than or equal to h, then x is merged with the cluster in which l belongs. Similarly, more than one clusters may be merged if each of them has a follower f' such that $||x - f'|| \leq h$.



(a) Spiral Dataset. (b) Speed-up factor of IncrementalSL.

Fig. 1. Performance of IncrementalSL on Spiral Dataset

IV. If distance between x and $l_{closest}$ is more than $1.5h$, x will be a new leader and creates a new cluster.

It can be shown that the clustering results produced by the IncrementalSL is exactly same as the results produced by the *al*-SL method after inserting all datapoints.

5 Performance Evaluation

In this section, we evaluate efficiency of the IncrementalSL method with one synthetic and one real world datasets. Plot of a synthetic data (Spiral Data) is shown in Fig. 1(a). As we discussed earlier that the IncrementalSL method produces same clustering results as produced by the *al*-SL method, therefore, we do not discusses about the clustering results produced by IncrementalSL in this section. We implemented IncrementalSL and *al*-SL method in C on Intel® Core^{TM} i5 Processor with 4 GB RAM Laptop. We use *speed-up factor*, which is defined as the ratio of the time taken by *al*-SL method and the time taken by our proposed IncrementalSL method. Experiments with Spiral dataset of size 9100 shows the IncrementalSL is upto one order faster (speed-up factor 9.89) than the *al*-SL method. To simulate incremental nature with this dataset, we added points with different sizes $(1820, 3600, 4500, 5500, 6000, 7000, 8000)$ in such a way that total points in the dataset does not exceed its size. It may be noted that inserting a set of points with size m indicates database had initially $(9100 - m)$ points. In the experiments, the *al*-SL was run on entire datasets while we ran IncrementalSL on incremental nature of the Spiral dataset. Plot in Fig. 1(b) shows the speed-factors of IncrementalSL over different sizes of inserted points.

We experimented with real Shuttle dataset (`http://archive.ics.uci.edu/ml/`), which has 9 integer-valued attributes of $58,000$ patterns distributed over 7 classes (after merging training and test sets). Class labels are eliminated from the all patterns. In experiments with Shuttle dataset, we inserted fixed size points $(3,000)$ to the

Table 1. Experiment with Shuttle Dataset

Size of Dataset	Time taken by IncrementalSL (in seconds)	Time taken by *al*-SL (in seconds)
5,000	1.09	1.32
10,000	2.31	3.41
20,000	5.11	8.15
30,000	8.21	14.31
40,000	11.01	23.21
58,000	16.21	36.21

different sizes datasets to simulate the incremental nature. It is found that the our proposed IncrementalSL is faster than the *al*-SL method. The detailed results are given in Table 1.

6 Conclusions and Future Research

The *al*-SL is a distance based clustering method, which is suitable for finding arbitrary shaped clusters in a static dataset. In this paper, we presented an incremental version of the *al*-SL method to work with dynamic scenarios. The clustering results produced by the proposed incremental method is exactly same as the results produced by *al*-SL method and the IncrementalSL is faster than the *al*-SL method. The IncrementalSL method utilizes the metric space properties to find affected region over insertion of data points. In the present work, we do not consider the effect of deletions of points in a dataset. Finding the clustering structures over deletions of points can be future research direction of this work. Another potential research direction can be the study of effect in clustering structures over change of the value of input parameter (h).

References

1. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proceedings of 2nd ACM SIGKDD, pp. 226–231 (1996)
2. Sneath, A., Sokal, P.H.: Numerical Taxonomy. Freeman, London (1973)
3. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognition Letters 31(8), 651–666 (2010)
4. Patra, B.K.: Mining Arbitrary Shaped Clusters in Large Dataset. PhD thesis, Indian Institute of Technology Guwahati, Guwahati, INDIA (2012)
5. Murty, M.N., Krishna, G.: A hybrid clustering procedure for concentric and chain-like clusters. Int. J. Comput. Inform. Sci. 10(6), 397–412 (1981)
6. Wong, M.A.: A hybrid clustering algorithm for identifying high density clusters. Journal of the American Statistical Association 77(380), 841–847 (1982)

7. Vijaya, P.A., Murty, M.N., Subramanian, D.K.: Efficient bottom-up hybrid hierarchical clustering techniques for protein sequence classification. Pattern Recognition 39(12), 2344–2355 (2006)
8. Lin, C.R., Chen, M.S.: Combining partitional and hierarchical algorithms for robust and efficient data clustering with cohesion self-merging. IEEE Trans. on Knowl. and Data Eng. 17(2), 145–159 (2005)
9. Chaoji, V., Hasan, M.A., Salem, S., Zaki, M.J.: Sparcl: an effective and efficient algorithm for mining arbitrary shape-based clusters. Knowl. Inf. Syst. 21(2), 201–229 (2009)
10. Patra, B.K., Nandi, S., Viswanath, P.: A distance based clustering method for arbitrary shaped clusters in large datasets. Pattern Recognition 44(12), 2862–2870 (2011)
11. Hartigan, J.A.: Clustering Algorithms. John Wiley & Sons, Inc., New York (1975)
12. Spath, H.: Cluster Analysis Algorithms for Data Reduction and Classification of Objects. Ellis Horwood, UK (1980)
13. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 3rd edn. Academic Press, Inc., Orlando (2006)
14. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: An efficient data clustering method for very large databases. In: Proceedings of ACM SIGMOD International Conference on Management of Data, SIGMOD 1996, pp. 103–114 (1996)
15. Charikar, M., Chekuri, C., Feder, T., Motwani, R.: Incremental clustering and dynamic information retrieval. In: Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing, pp. 626–635 (1997)
16. Chen, C.-Y., Hwang, S.-C., Oyang, Y.-J.: An incremental hierarchical data clustering algorithm based on gravity theory. In: Chen, M.-S., Yu, P.S., Liu, B. (eds.) PAKDD 2002. LNCS (LNAI), vol. 2336, pp. 237–250. Springer, Heidelberg (2002)
17. Widyantoro, D., Ioerger, T., Yen, J.: An incremental approach to building a cluster hierarchy. In: Proceedings of IEEE International Conference on Data Mining, ICDM 2003, pp. 705–708 (2002)
18. Ester, M., Kriegel, H.P., Sander, J., Wimmer, M., Xu, X.: Incremental clustering for mining in a data warehousing environment. In: Proceedings of 24th International Conference on Very Large Data Bases (VLDB 1998), pp. 323–333 (1998)

C-MAP: Framework for Multi-agent Planning in Cyber Physical Systems

Sumant Mukherjee[1] and Santanu Chaudhury[2]

[1] Institute for Systems Studies & Analysis, DRDO, Delhi, India
{sumantiitd,schaudhury}@gmail.com
[2] Dept. of Electrical Engineering, IIT, Delhi, India

Abstract. In this paper we propose a novel multi-agent planning framework that attempts to deal with dynamic nature of cyber physical system while planning and show usefulness of our approach in exploring different strategies and configurations of planning architecture. The proposed approach use multi-agent planners at different layers of abstraction to collaborate by using concepts of task sharing, brokering of task allocation, continuous adaptive planning and blackboard based multi-agent interaction.

Keywords: cyber-physical system, multi-agent planning, task network, contract net protocol, blackboard.

1 Introduction

Cyber physical system (CPS) [1] intensively interact with physical processes where sensors sense the physical world, computational components process sensor data and activate actuators to influence the physical process in real time. The computational components for CPS need to be adaptive to deal with uncertainty of observations from sensor and outcome of actuator actions. Information-based warfare [2] is the term used in military circles to define network centric battle space that can be visualized as military CPS consisting of network of sensors, weapons and Command & Control (C2) entities. Here the plans from multiple C2 planners need to be coordinated using various techniques of multi-agent planning (MAP) [3]. MAP for military CPS is an evolving technology for intelligent collaboration between multiple planner agents. Traditional decision-theoretic models become intractable when entities in CPS grow or the uncertainty about the environment increases.

This paper provides a novel MAP framework to deal with dynamic nature of CPS while planning using market-based algorithms [4], biologically inspired continuous planning [5] and blackboard-based [6] coordination between the planning layers. Domain of combat battlefield is used as case study to implement market-based algorithm for top-down task allocation to lower level planners. Bottom level planners use biologically inspired continuous planning to dynamically adapt to changing battlefield environment. Coordination between the planning layers is done using Blackboard. The effectiveness of the proposed approach has been illustrated using simulation testbed that generates typical military combat scenario.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 237–242, 2013.

Section 2 provides an overview of proposed MAP framework and its components. Proposed approach along with application scope of market-based algorithm, continuous planning and blackboard-based coordination in MAP is described in section 3. Section 4 shows the simulation experiments to illustrate the effectiveness of the proposed approach. Concluding remarks with future scope of the study are given in the last section.

2 MAP for CPS: An Overview

Planning problem in military CPS environment is a complex distributed problem solving task as it needs to harness inherently distributed resources and require communication and coordination between multiple C2 planners. We propose to deal with the complexity of distributed planning process in CPS environment using layers of abstraction for strategic, operational and tactical planning and collaboration between different layers of multi-agent planners. To construct a collective plan for working together, our higher level planning agents (strategic and operational planner) decompose task into sub-task (task decomposition), allocate sub-task to lower level planning agents (task allocation to operational and tactical planners), exchange sub-problem solution (result sharing) and synthesize solution using Blackboard (see Fig 1).

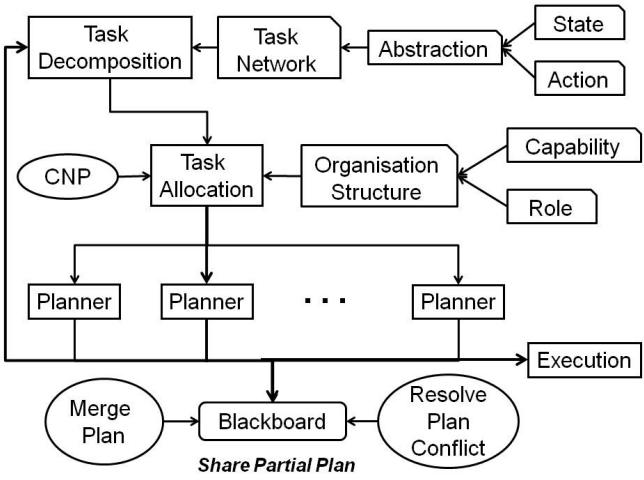


Fig. 1. Multi-Agent Planning Architecture

3 Proposed Approach

The paper proposes a unique MAP approach based on top-down concept of market-based algorithms, bottom-up biologically inspired continuous planning and Blackboard-based coordination between the planning layers. Proposed Combat Multi-Agent Planner (C-MAP) implements layers of planners that reflect hierarchical organization

of military (see Fig. 2). Global Commander (GC) is top level C2 layer that coordinates actions of the Local Commander (LC) which in turn coordinates actions of lowest level Unit commander (UC) under its command. C-MAP carries out major military tasks of surveillance, defensive and offensive operations in the simulated battlespace on need basis. For each operations GC generates set of operational tasks that all LC bids for and on allocation further decomposes the allocated operational task into tactical tasks. Task decomposition is carried out by making use of domain task network that consists of structured representation of set of tasks and inter-relationship between them based on state and action abstraction. The domain task network encode methods to achieving goals of surveillance, defense and offense by using operators to maneuver and engage threats, maintaining force balance, massing of units and helping peers in trouble.

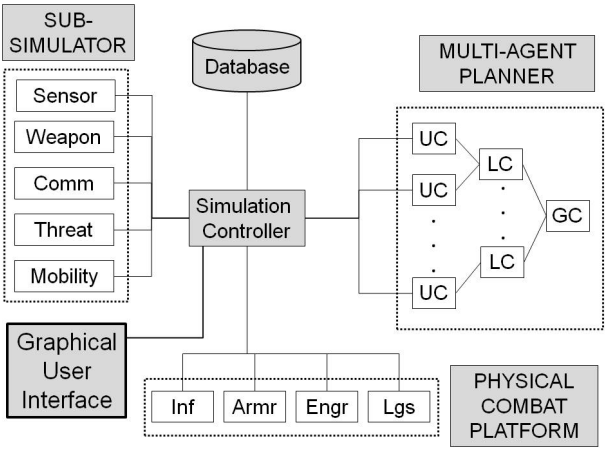


Fig. 2. Testbed Architecture

Market-based approach of contract net protocol (CNP) is used for task allocation in top-down fashion. Participants (LC or UC) bid for task depending on capability, role, current load of agents and task completion deadlines. Bidders use formally encoded multi-agent organizational structure that associates with each agent the types of tasks it can carry out and its priority over such tasks (this facilitates overlapping of agent capabilities). C-MAP multi-agent organizational structure encodes capabilities required to carry out a particular task, capabilities of particular type of agent and its prioritization over tasks

After successful task allocations, agents carry out adaptive continuous planning in bottom-up fashion using biologically inspired concept [5]. Combat is modeled as complex adaptive system where combat forces are visualized to be composed of large numbers of nonlinearly interacting parts. C-MAP agents continually adapt to changing combat environment and specifically adopt different "tendency" such as cluster with healthy friendly units, support injured friendly units, preserve self by avoiding healthy enemy units, be aggressive to injured enemy units and so on.

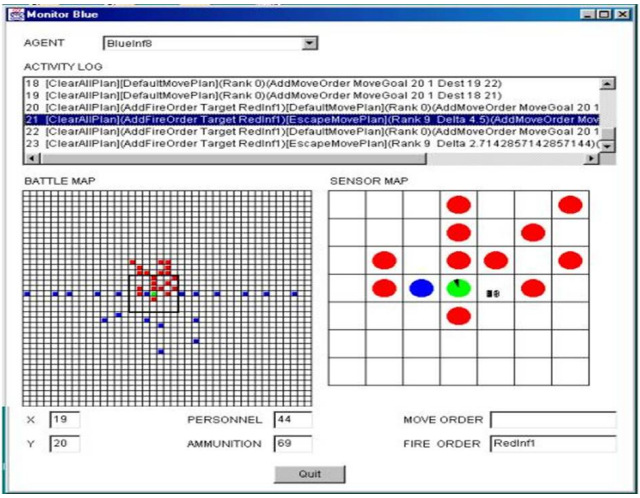


Fig. 3. C-MAP User Interface

Coordination between different layers of planners is achieved using Blackboard based inter-agent interaction. Different planner agents add structured knowledge tuple to Blackboard which others can query, update or delete. C-MAP Blackboard-Control agent directs the coordination process by managing how knowledge source (KS) agent responds to relevant event. KS agents act as critic of plans from different perspectives of task sharing opportunities, fusion of information, temporal consistency, and constraint violation. The task sharing critic agent looks for opportunities of sharing of common tasks. Possible opportunities could be achieving surveillance task as side-effect of offensive or defensive task, supplying of resources or clearing obstacles in proximity and so on. The task sharing critic agent identifies these opportunities by monitoring task related tuple such as agent active task assignment, task capabilities, agent capability and triggering task sharing event when matching rules fire. The fusion critic agent fuses lower level information to higher abstraction such as size of enemy-deployment (based on collocation of enemy units) or prediction of move-vector of enemy-units (based on temporal tracking of enemy unit locations) [7]. The temporal critic agent monitor task with respect to its deadlines (gap between contracted task and actual state at end of contract) and trigger appropriate corrective actions such as additional operational or tactical tasks. The consistency critic agent checks for different constraint violation on tasks such as resource constraint such as unit does not have sufficient fuel or ammunition to carry out particular task. If a constraint violation is detected then consistency critic agent send "constraint-violation" message to relevant unit indicating that corrective action needs to be taken.

4 Simulation Results

We have implemented a combat simulation platform (CSP) to illustrate effectiveness of the proposed approach (see Fig. 2). CSP simulates typical military combat scenario

in dynamic military CPS environment constituting of physical combat systems (PCS) with heterogeneous capability. Combat scenario gives details of total force composition of both blue and red side, their locations and resources such as sensors, weapons, communication devices, personnel, ammunition and fuel. Combat environment consists of terrain, weather and day/night characteristics where weather parameters effects sensor probabilities of detection Pd, Weapon probability of kill P_k and communication reliability. CSP provides PCS with limited local sensor information with uncertainty associated to it and restricted communication capability.

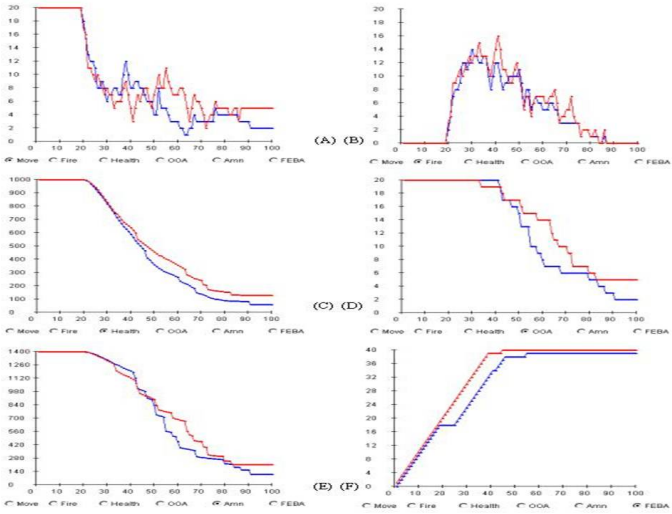


Fig. 4. Combat Action analysis curves

CSP constitutes of simulation controller (SC), sub-simulators and database module. C-MAP agents connect to respective PCS in CSP via SC. SC sends initial setting of the battlespace to each C-MAP agent. During simulation loop SC sends corresponding PCS sensor information to each agent and each agent can submit an action command to the SC individually. The SC sends agent action commands to sub-simulators (combat domain models for movement, threat, sensor, weapon, communication, obstacles, environment and attrition). These sub-simulators incorporate their respective effects and update the state of the battlespace. SC sends updated battlespace states to the GUI (see Fig. 3) and advances the simulation clock.

C-MAP can be configured to be completely decentralized (where units take independent decision without any C2 layer to guide them) or centralized C2 (i.e. units are guided by Global as well as local HQ). Fig. 4 depicts a simulation run output in term of combat actions taken by C-MAP agents of blue and red sides. The combat action analysis curves indicate effects of different configuration of C-MAP parameters on combat actions of maneuverability of units, casualties suffered, expenditure of resources etc. We believe that combination of CSP testbed and pluggable C-MAP multi-agent planning framework can be a useful tool for exploring different strategies for configurations for MAP architecture in CPS environment.

5 Conclusion

In this paper we have proposed C-MAP, a novel multi-agent planning architecture to deal with dynamic nature of cyber-physical worlds where underlying planner adapts to dynamic initial state, goal state and planning domain. C-MAP use different abstraction layers of planning and use concepts of task network, CNP based negotiation, Blackboard-based agent interaction. The contribution of this paper lies in providing a framework to explore interplay between top-down marketplace dynamics and bottom-up biologically inspired adaptive continuous planning to deal with dynamic CPS environment. To explore efficacy of C-MAP framework we have also implemented a testbed that simulates dynamic military CPS environment constituting of physical systems with heterogeneous capability. In conclusion it is worth observing that proposed framework is the first (to the best of our knowledge) that uses above ideas in a single framework. Our future work will focus on developing mechanism of semantic mediation to overcome heterogeneity challenges in CPS at different abstractions of physical devices, data/information and services using ontology model.

References

1. Tan, Y., Goddard, S.: A prototype architecture for cyber-physical systems. SIGBED Rev. 5(1), 1–2 (2008)
2. Alberts, D.S., et al.: Network Centric Warfare: Developing and Leveraging Information Superiority, 2nd edn. CCRP Press
3. Durfee, E.H.: Distributed Problem Solving and Planning. In: Weiss, G. (ed.), pp. 121–164. MIT Press, Cambridge (1999)
4. Smith, R.G.: The Contract Net Protocol. IEEE Trans. on Computers C-29(12) (December 1980)
5. Ilachinski, A.: Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Warfare. Center for Naval Analyses Research Memorandum CRM 97-61 (June 1997)
6. Corkill, D.D.: Collaborating Software Blackboard and Multi-Agent Systems & the Future. In: Proceedings of the International Lisp Conference (2003)
7. Gupta, M., Mukherjee, S.: Towards Situation Awareness in Integrated Air Defence Using Clustering and Case Based Reasoning. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) PReMI 2009. LNCS, vol. 5909, pp. 579–584. Springer, Heidelberg (2009)

A New Intelligent Approach for Mobile Robot Navigation

Prases Kumar Mohanty* and Dayal R. Parhi

Robotics Laboratory, National Institute of Technology, Rourkela 769008, Odisha, India
pkmohanty30@gmail.com,
dayalparhi@yahoo.com

Abstract. In recent times computational intelligent techniques such as fuzzy inference system (FIS), artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) are mainly considered as effective and suitable optimization methods for modeling an engineering system. In this paper an efficient hybrid technique has been applied for mobile robot navigation using multiple adaptive neuro-fuzzy inference system (MANFIS). ANFIS has taken the advantages of both fuzzy inference system and artificial neural network. First, we design an adaptive fuzzy controller with four input parameters, two types of output parameters and three parameters each. Next each adaptive fuzzy controller acts as a single Sugeno-Takagi type fuzzy inference system where inputs are the different sensor based information and output corresponds to the velocity of the mobile robot. The implementation of the proposed navigational controller is discussed via numerous simulation examples. It is found that such an adaptive neuro-fuzzy controller is successfully and quickly finding targets in an unknown or partially unknown environment.

Keywords: Mobile robot, Navigation, Neuro-fuzzy, Obstacle avoidance.

1 Introduction

At present mobile robots have been successfully used in various areas of engineering such as aerospace research, nuclear research, production engineering etc. The major objective in the current robotic research field is to find a collision free path from a given start position to predefined destination point. In general path planning algorithms are classified as local and global depending upon the surrounding environment. In global path planning the surrounding environment is completely known to the mobile robot so the path travelled by the mobile robot is predefined, where as in local path planning the environment is completely unknown or partially known to the mobile robot so various sensors are used to perceive the information about the surrounding environment and plan the motion accordingly. Many exertions have been paid in the past to improve various robot navigation algorithms.

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 243–249, 2013.

In literature survey, there can be found several researchers have been addressed on many intelligent techniques for mobile robot navigation. Many authors have considered a controller with complete information of the environment [1-2].Due to the complexity and uncertainty of the path planning problem, classical path planning methods, such as Visibility Graph [3], Voronoi diagrams [4], Grids [5], Cell decomposition [6], artificial potential field [7], Rule based methods [8], and Rules learning techniques [9] are not appropriate for path planning in dynamic environments. The use of the above algorithms for path finding for mobile robot requires more time and the finding of this path will not completely feasible for real-time movement. There are many fuzzy logic methods using various implementations or in combination with other techniques [10-14]. Mobile robot path planning based on neural network approaches presented by many researchers [15-18].Among the intelligent techniques ANFIS is a hybrid model which combines the adaptability capability of artificial neural network and knowledge representation of fuzzy inference system [19].Song and Sheen [20] developed a pattern recognition method based on fuzzy-neuro network for reactive navigation of a car-like robot. Li et al. [21] suggested a neuro-fuzzy technique for behavior based control of a car-like robot that navigates among static obstacles. Navigation of multiple mobile robots using Neuro-fuzzy technique addressed by Pradhan et al. [22]. In this design, output from the neural network given as input to the fuzzy controller to navigate the mobile robot successfully in the clutter environment. Experimental verifications also have been done with the simulation results to prove the validity of the developed technique. Navigation of mobile robots using adaptive neural-fuzzy system discussed by Nefti et al. [23]. Different sensor based information they have given to the SugenoTakagi fuzzy controller and output from the controller is the robot orientation. Experimental results settle the importance of the methodology when dealing with navigation of a mobile robot in unknown or partially unknown environment. A Neuro-Fuzzy Controller based mobile robot navigation presented by Kim and Trivedi [24]. In this study they have implemented neural integrated fuzzy controller to control the mobile robot motion in terms of steering angle, heading direction, and speed. To determine collision-free path of mobile robot navigating in a dynamic environment using Neuro-fuzzy technique presented by Hui et al. [25]. In this paper the performances of Neuro-fuzzy approaches are compared with other approaches (GA, Mamdani) and it was found that Neuro-fuzzy approaches are found to perform better than the other approaches. Control of mobile robot based on Neuro-fuzzy technique discussed by Godjevac and Steele [26]. In this paper they have shown how Neuro-fuzzy controllers can be achieved using a controller based on the Takagi-Sugeno design and a radial basis function neural network for its implementation.

We propose in this paper to develop an intelligent navigational controller for solving navigation problem for mobile robot in an unknown or partially unknown environment. A new MANFIS (Multiple Adaptive Neuro-Fuzzy Inference System) controller has been designed to solve the optimization problem. Finally,

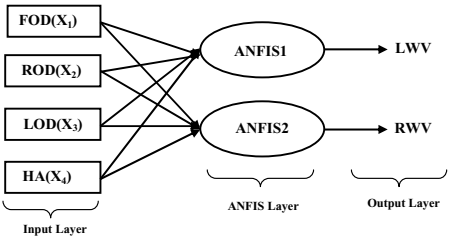


Fig. 1. Multiple ANFIS (MANFIS) Controller for Mobile Robot Navigation

simulation results are presented to verify the effectiveness of the proposed controller in various scenarios populated by stationary obstacles.

2 Architecture of Multiple Adaptive Neuro-Fuzzy Inference System (MANFIS) for Current Analysis

Adaptive network-based fuzzy inference system (ANFIS) is one of hybrid intelligent neuro-fuzzy system and it functioning under Takagi-Sugeno-type FIS, which was developed by Jang [19] in 1993. There are two learning techniques are used in ANFIS to show the mapping between input and output data and to compute optimized of fuzzy membership functions. These learning methods are back propagation and hybrid. Parameters associated with fuzzy membership functions will modify through the learning process.

As for the prediction of left wheel velocity (LWV) and right wheel velocity (RWV) for mobile robot we assume each adaptive neuro-fuzzy controller under consideration of four inputs i.e. Front obstacle distance(FOD) (x_1), Right obstacle distance(ROD) (x_2), Left obstacle distance(LOD)(x_3), Heading angle(x_4) (TA) and each input variable has three bell membership functions(MF) A_1 (Far), A_2 (Medium) and A_3 (Near) , B_1 (Far), B_2 (Medium) and B_3 (Near), C_1 (Far), C_2 (Medium) and C_3 (Near), D_1 (Negative), D_2 (Zero) and D_3 (positive) respectively, then a Takagi-Sugeno-type fuzzy inference system if-then rules are set up as follows;

Rule: if x_1 is A_i and x_2 is B_i and x_3 is C_i and x_4 is D_i, then

f_n(wheel velocity) = $p_n x_1 + q_n x_2 + r_n x_3 + s_n x_4 + u_n$

where, i=1,2,3 and p_n, q_n, r_n, s_n and u_n are the linear parameters of function f_n and changing these parameters we can modify the output of ANFIS structure. The function of each layer in ANFIS model is discussed as follows:

Input Layer : In this layer nodes simply pass the incoming signal to layer-1. That is

$$O_{0,FOD} = x_1, O_{0,ROD} = x_2, O_{0,LOD} = x_3, O_{0,HA} = x_4 \tag{2.1}$$

First Layer: This layer is the fuzzification layer. Neurons in this layer complete fuzzification process. Every node in this stage is an adaptive node and

calculating the membership function value in fuzzy set. The output of nodes in this layer are presented as

$$O_{1,i} = \mu_{A_i}(x_1), O_{1,i} = \mu_{B_i}(x_2), O_{1,i} = \mu_{C_i}(x_3), O_{1,i} = \mu_{D_i}(x_4) \tag{2.2}$$

i=1,2,3

Here O_1,i is the bell shape membership grade of a fuzzy set S ($A_i, B_i, C_i$and D_i) and it computing the degree to which the given inputs ($x_1, x_2, x_3$and x_4) satisfies the quantifier S. Membership functions defined as follows;

$$\mu_{A_i}(x) = \frac{1}{1 + [(\frac{x_1 - c_i}{a_i})^2]^{b_i}}, \tag{2.2a}$$

$$\mu_{B_i}(x) = \frac{1}{1 + [(\frac{x_2 - c_i}{a_i})^2]^{b_i}} \tag{2.2b}$$

$$\mu_{C_i}(x) = \frac{1}{1 + [(\frac{x_3 - c_i}{a_i})^2]^{b_i}} \tag{2.2c}$$

$$\mu_{D_i}(x) = \frac{1}{1 + [(\frac{x_4 - c_i}{a_i})^2]^{b_i}} \tag{2.2d}$$

$a_i, b_i$and c_i are parameters that control the Centre, width and slope of the Bell-shaped function of node i respectively. These are also known as premise parameters.

Second Layer: It is also known as rule layer. Every node in this layer is a fixed node and labeled as π_n. Every node in this stage corresponds to a single Sugeno-Takagi fuzzy rule. Each rule point receives inputs from the respective points of layer-2 and calculates the firing strength of the each fuzzy rule. Output from each node is the product of all incoming signals.

$$O_{2,n} = W_n = \mu_{A_i}(x_1).\mu_{B_i}(x_2).\mu_{C_i}(x_3).\mu_{D_i}(x_4) \tag{2.3}$$

Where W_n represents the firing strength or the truth value, of nth rule and n=1, 2, 3.....81 is the number of Sugeno-Takagi fuzzy rules.

Third Layer: It is the normalization layer. Every node in this layer is a fixed node and labeled as N_n. Each point in this layer receives inputs from all points in the adaptive fuzzy rule layer and calculates the normalized firing strength of a given rule. The normalized firing strength of the nth point of the nth rules firing strength to sum of all ruless firing strength.

$$O_{3,n} = \overline{W}_n = \frac{W_n}{\sum_{n=1}^{81} W_n} \tag{2.4}$$

The number of points in this layer is the same the number of points in the fuzzy layer that is 81 points. The output of this layer is called normalized firing strength.

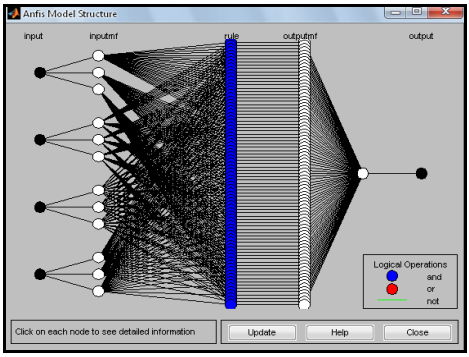


Fig. 2. The structure of ANFIS 1 network

Fourth Layer: Every node in this layer is an adaptive node. Each node in this layer is connected to the corresponding normalization node, and also receives initial inputs x_1, x_2, x_3 and x_4. A defuzzification node determines the weighted consequent value of a given rule define as,

$$O_{4,n} = \overline{W}_n f_n = \overline{W}_n [p_n(x_1) + q_n(x_2) + r_n(x_3) + s_n(x_4) + u_n] \tag{2.5}$$

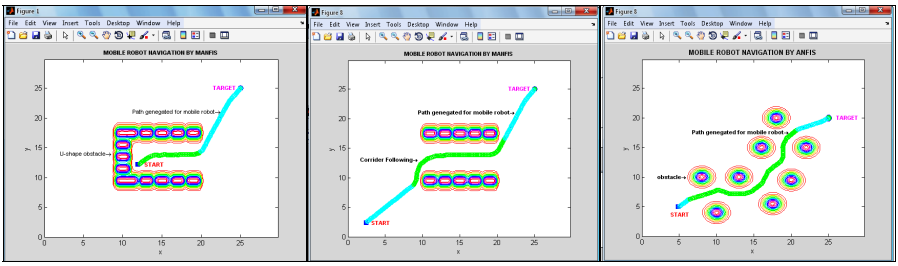
Where $\overline{W}_n$ is a normalized firing strength from layer-3 and are the parameters set of this node. These parameters are also called consequent parameters.

Fifth Layer: It is represented by a single summation node. This single point is a fixed point and labeled as $\sum$. This point determines the sum of outputs of all defuzzification points and gives the overall model output that is velocity of wheels.

$$O_{5,n} = \sum_{n=1}^{81} \overline{W}_n f_n = \frac{\sum_{n=1}^{81} W_n f_n}{\sum_{n=1}^{81} W_n} \tag{2.6}$$

3 Simulation Results and Discussion

In this part MATLAB simulations of the proposed algorithm are presented. In order to verify the validity and performance of the current navigational controller, we have performed the simulation results to verify robot trajectory under the various environmental scenarios according to the robot motion rule. When a robot is close to an obstacle, it must change its speed to avoid the obstacle shown by green color line in simulation figure. If a target is sensed by a mobile robot, it will decide whether it can reach that target, i.e. it will judge whether there are obstacles that will obstruct its path. If the path leading to the target is clear, the robot will turn and proceed towards the target shown by cyan line. In Fig. 3a-3c shows the path created for mobile robot motion in various environments with considering different start and goal positions. It was clearly



(a) Single mobile robot escaping from dead end (b) Single mobile robot escaping from corridor (c) Single mobile robot in maze environment

Fig. 3. Mobile robot navigation using MANFIS controller

observed, through MATLAB simulations, that the proposed hybrid controller technique performs favorably over each individual reactive behavior. By using this technique the mobile robot can reach successfully at the target object.

4 Conclusion

It has been shown that the proposed navigational controller can be successfully implemented for a single target with unknown or partially unknown environment .The main emphasis of the model was placed on flexibility in adaption to various complex environments and robustness to perception uncertainty. The obtained results from above technique were analyzed in a number of simulated experiments and it was clearly observed that the current navigational controller is an effective approach for the obstacle avoidance and moving towards the goal by the mobile robot. Future work involves extending this research work to multiple mobile robots with dynamic obstacles instead of single robot with static obstacles.

References

1. Latombe, J.C.: Robot Motion Planning. Kluwer Academic Publishers, New York (1990)
2. Canny, J.E.: The Complexity of Robot Motion Planning. MIT Press, Cambridge (1988)
3. Lozano-Perez, T.: A simple motion planning algorithm for general robot manipulators. IEEE Journal of Robotics and Automation 3, 224–238 (1987)
4. Leven, D., Sharir, M.: Planning a purely translational motion for a convex object in two dimensional space using generalized voronoi diagrams. Discrete & Computational Geometry 2, 9–31 (1987)
5. Payton, D., Rosenblatt, J., Keirsey, D.: Grid- based mapping for autonomous mobile robot. Robotics and Autonomous Systems 11, 13–21 (1993)
6. Regli, L.: Robot Lab: Robot Path Planning. Lectures Notes of Department of Computer Science. Drexel University (2007)

7. Khatib, O.: Real time Obstacle Avoidance for manipulators and Mobile Robots. In: IEEE Conference on Robotics and Automation, vol. 2, p. 505 (1985)
8. Fujimura, K.: Motion Planning in Neritic Environments. Springer (1991)
9. Ibrahim, M.Y., Fernandes, A.: Study on Mobile Robot Navigation Techniques. In: IEEE International Conference on Industrial Technology, December 8-10, vol. 1, pp. 230–236 (2004)
10. Huq, R., Mann, G.K.I., Gosine, R.G.: Mobile robot navigation using motor schema and fuzzy content behavior modulation. Application of Soft Computing 8, 422–436 (2008)
11. Selekwa, M.F., Dunlap, D.D., Shi, D., Collins Jr., E.G.: Robot navigation in very cluttered environment by preference based fuzzy behaviors. Autonomous System 56, 231–246 (2007)
12. Abdessemed, F., Benmahammed, K., Monacelli, E.: A fuzzy based reactive controller for a non-holonomic mobile robot. Robotics Autonomous System 47, 31–46 (2004)
13. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Fuzzy logic techniques for navigation of several mobile robots. Application of Soft Computing 9, 290–304 (2009)
14. Motlagh, O., Tang, S.H., Ismail, N.: Development of a new minimum avoidance system for behavior based mobile robot. Fuzzy Sets System 160, 1929–1946 (2009)
15. Velagic, J., Osmic, N., Lacevic, B.: Neural Network Controller for Mobile Robot Motion Control. World Academy of Science. Engineering and Technology 47, 193–198 (2008)
16. Singh, M.K., Parhi, D.R.: Intelligent Neuro-Controller for Navigation of Mobile Robot. In: Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai, Maharashtra, India, pp. 123–128 (2009)
17. Castro, V., Neira, J.P., Rueda, C.L., Villamizar, J.C., Angel, L.: Autonomous Navigation Strategies for Mobile Robots using a Probabilistic Neural Network (PNN). In: 33rd Annual Conference of the IEEE Industrial Electronics Society, Taipei, Taiwan, pp. 2795–2800 (2007)
18. Yang, S.X., Meng, M.: An efficient neural network approach to dynamic robot motion planning. Neural Network 13, 143–148 (2000)
19. Jang, J.S.R.: ANFIS: Adaptive network-based fuzzy inference system. IEEE Transaction on System, Man and Cybernetics Part B 23, 665–685 (1993)
20. Song, K.T., Sheen, L.H.: Heuristic fuzzy-neuro network and its application to reactive navigation of a mobile robot. Fuzzy Sets Syst. 110, 331–340 (2000)
21. Li, W., Ma, C., Wahl, F.M.: A neuro-fuzzy system architecture for behavior based control of a mobile robot in unknown environments. Fuzzy Sets Syst. 87, 133–140 (1997)
22. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Neuro-fuzzy technique for navigation of multiple mobile robots. Fuzzy Optimum Decision Making 5, 255–288 (2006)
23. Nefti, S., Oussalah, M., Djouani, K., Pontnau, J.: Intelligent Adaptive Mobile Robot Navigation. Journal of Intelligent and Robotic Systems 30, 311–329 (2001)
24. Ng, K.C., Trivedi, M.M.: A Neuro-Fuzzy Controller for Mobile Robot Navigation and Multi robot Convoying. IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics 28, 829–840 (1998)
25. Hui, N.B., Mahendar, V., Pratihar, D.K.: Time-optimal, collision-free navigation of a car-like mobile robot using neuro-fuzzy approaches. Fuzzy Sets and Systems 157, 2171–2204 (2008)
26. Godjevac, J., Steele, N.: Neuro-fuzzy control of a mobile robot. Neuro-Computing 28, 127–142 (1999)

Rapid Game Strategy Evaluation Using Fuzzy Extreme Learning Machine

YingJie Li, Peter Hiu Fung Ng, and Simon Chi Keung Shiu

Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract. Interactions among game units can be conveniently described by fuzzy measures and integrals. Focusing on Warcraft, there are several good results of unit selection strategy evaluation for a genetic algorithm that search in plan space. However, this kind of evaluators are suffered from high complexity in fuzzy measure determination. In this paper, we novelly combine Extreme Learning Machine(ELM) and Fuzzy Integral(FI) to achieve a fast evaluation of game strategy. Experimental comparison demonstrates the effectiveness of the proposed method in both time and accuracy.

1 Introduction

Real time strategy(RTS) games is a genre of computer war-games which does not process incrementally in turn. The core of this game is to build up an army with appropriate unit combination which can gains massive destroy power in a short time. Specially, feature interactions are proofed existing among different unit types [1,2], which offer a large variety of fundamental AI research problems such as decision making under uncertainty[3,4] and adversarial real-time planning [5].In this paper, we concentrate on how to evaluate different unit combinations rapidly and a combination of ELM and FI with real game data is introduced .

Our technical contributions are:

- A new fuzzy integral named mean-based fuzzy integral (Me-based FI) is developed to calculate the intergraded power of different unit combinations. It allows us to capture all related unit types and average the related effectiveness accurately.
- A novel structure of ELM is developed to learn CI and Me-based FI. In particular, we introduce a novel set selection mechanism to reflect the complex nonlinear relation in the model.

In section 2, we briefly introduce the fuzzy measure and integral. Section 3 details the Fuzzy Extreme Learning Machine(FELM) for unit combination strategy evaluation. We then provide detailed experimental analysis of our technique in section 4. Finally, we have the conclusion in section 5.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 250–255, 2013.

2 Basic Concepts

2.1 Fuzzy Measure

In order to overcome the limitations of classical measure, fuzzy measure replacing the additive property with weaker conditions as monotonicity [6]. The definition of fuzzy measure is shown as follows:

Definition 1. *Let (X, F) be a measurable space. A fuzzy measure is a real-valued set function $\mu : F \to (+\infty, -\infty)$ that satisfies the following criteria:*
(MM1) $\mu(\emptyset) = 0$;
(MM2) $\mu(A) \geq 0$ for every $A \in F$;
(MM3) $\mu(A) \geq \mu(B)$ whenever $A \in F, B \in F, A \subseteq B$.

Set function μ is regarded as a signed efficiency measure if it only satisfies the condition (MM1)[7]. It has the greatest representation ability but also the highest complexity. According to the definition, one requires to defined $2^n - 1$ coefficients for n variables. Considering the demand of high accuracy in game, we apply it to describe the interactions among units. Each coefficient represents the combat effectiveness of corresponding unit combination.

2.2 Fuzzy Integral

Fuzzy integral is developed to integrating information with respect to fuzzy measure. Several types of fuzzy integrals have been developed for different applications [8]. Choquet integral,as 1, is the most widely used one that has attracted significant interest and success. The definition is shown as follows:

Definition 2. *Given a fuzzy measure μ on x. The discrete Choquet integral of a function $f : X \to \Re$ can be defined as:*

$$(c) \int f(x) \cdot \mu(X) = \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] \cdot \mu\{x \mid sf(x) \geq f(x_i)\}, \qquad (1)$$

where $f(x_0) = 0$ and $f(x_0) \leq f(x_1) \leq f(x_2) \leq \ldots \leq f(x_n)$.

3 Methodology

3.1 Strategy Definition in RTS Games

Strategy Case={Goal, Situation, Score}

Strategy is defined as arranging the army with appropriate unit selection which able to gain massive destroy power against particular enemy. "Goal" is the creation of suitable unit combination with certain proportion, e.g. 10% peasants, 30% footman and 60% rifleman. "Situation" is defined as what circumstances the player is dealing with. Cases are clustered into several groups based on different races and unit combinations. "Score" is the evaluation point provided by game system, which are adopted in learning the fuzzy measures.

3.2 Mean-Based Fuzzy Integral

Me-based FI, as 2, is designed to release the problem of interactions selection by involving all the fuzzy measure with the current unit type. It performs a better result than CI but suffers from high complexity of set selection by using GA based technique. In this research, we overcome this shortage by combining it with ELM.

Definition 3. *Given a fuzzy measure μ on x. The discrete Mean-based Fuzzy Integral of a function $f : X \to \Re$ can be defined as:*

$$(m)\int f(x) \circ \mu(X) = \sum_{i=1}^{n} x_i \times (\frac{1}{m_i} \sum_{j=1}^{m_i} \mu(S_{ij})) \tag{2}$$

where $x_i \in S_{ij}$ and $\forall x \in S_{ij}$,$x \neq 0$, n is number of unit type, m is the number of set which include x_i.

3.3 Design of ELM

The essence of ELM is to randomly assign the input biases, while analytically calculate the output weights by pseudo inverse of the hidden layer output matrix[9]. In this research, we aim to design the ELM model based on the equivalent forms of fuzzy integrals. With these formulas, the value of CI and Me-based FI could be expressed as a linear function that explains the relation between f value and the corresponding fuzzy measure respectively. The designed FELM model is show as 1.

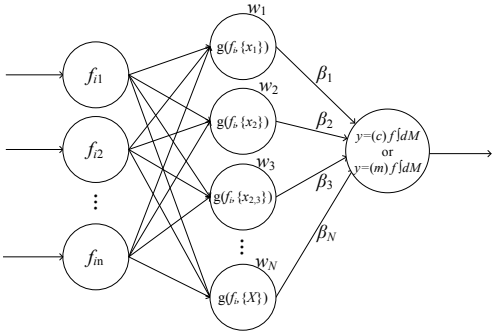


Fig. 1. Designed ELM describing fuzzy integrals

Generally, the learning process of the ELM algorithm contains two stages as follows, i.e., set selection stage and learning stage.

Stage 1: Set selection

In this stage, the input data will be transferred to another form based on Wang's selection algorithm[10]. The purpose is to represent the set selection process of fuzzy integrals with different information source. The set selection algorithm of CI and ME-based FI is shown as 3, 4 respectively.

$$g_{ij} = \max(\min_{k_j \in K_j} f^{(i)}(x_{k_j}) - \max_{k_j \in \overline{K_j}} f^{(i)}(x_{k_j}), 0), \tag{3}$$

$$g_{ij} = \underset{k_j \in K_j}{\text{avg}} f(x_{k_j}) \tag{4}$$

where$K_j = \left\{k : \frac{j}{2^k} - \left\lfloor \frac{j}{2^k} \right\rfloor \geq 0.5,\ 1 \leq k \leq n\right\}$, $\overline{K_j} = \{1, 2, \cdots, n\} - K_j$, $\min_{k\in\emptyset} f_k = \max_{k\in\emptyset} f_k = 0$,$i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, 2^n - 1$, m is the number of samples, n is the number of features, and $\lfloor a \rfloor$ denotes the integer part of the non-negative number α. Then, the $m \times n$ sample matrix can be transferred to a $m \times 2^n - 1$ input matrix as follows:

$$\mathrm{g} = \begin{bmatrix} g(f_1,\{x_1\}) & g(f_1,\{x_2\}) & g(f_1,\{x_1x,_2\}) & \ldots & g(f_1,X) \\ g(f_2,\{x_1\}) & g(f_2,\{x_2\}) & g(f_2,\{x_1x,_2\}) & \ldots & g(f_2,X) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g(f_m,\{x_1\}) & g(f_m,\{x_2\}) & g(f_m,\{x_1x,_2\}) & \ldots & g(f_m,X) \end{bmatrix}_{m\times(2^n-1)}.$$

Stage 2: Learning the fuzzy measure

Step 1: Randomly assign input weight w_{ij}, where $i = 1, \ldots, m$ and $j = 1, \ldots, 2^n - 1$;

Step 2: Calculate the hidden layer output matrix H as (5):

$$H = \begin{bmatrix} w_1 \cdot g(f_1,\{x_1\}) & w_2 \cdot g(f_1,\{x_2\}) & w_3 \cdot g(f_1,\{x_1,x_2\}) & \ldots & w_N \cdot g(f_1,\{x_A\}) \\ w_1 \cdot g(f_2,\{x_1\}) & w_2 \cdot g(f_2,\{x_2\}) & w_3 \cdot g(f_2,\{x_1,x_2\}) & \ldots & w_N \cdot g(f_2,\{x_A\}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_1 \cdot g(f_m,\{x_1\}) & w_2 \cdot g(f_m,\{x_2\}) & w_3 \cdot g(f_m,\{x_1,x_2\}) & \ldots & w_N \cdot g(f_m,\{x_A\}) \end{bmatrix}; \tag{5}$$

Step 3: Calculate the output weight vector $\beta = H^{\dagger}T$. $H^{\dagger}$ is the Moore-Penrose generalized inverse of matrix H. T is the output vector with m elements. Each element is the performance score of corresponding replay file.

Step 4: The learned fuzzy measure value could be calculated as (6):

$$\begin{bmatrix} \mu(\{x_1\}) \\ \mu(\{x_2\}) \\ \mu(\{x_1,x_2\}) \\ \vdots \\ \mu(X) \end{bmatrix} = \begin{bmatrix} w_1 \cdot \beta_1 \\ w_2 \cdot \beta_2 \\ w_3 \cdot \beta_3 \\ \vdots \\ w_{2^n-1} \cdot \beta_{2^n-1} \end{bmatrix}. \tag{6}$$

4 Experiment

4.1 Data Collection and Preprocessing

We select the RTS game named Warcraft III, a famous RTS game with over 7 million copies sold, as our research platform. 2,649 game files of professional one-versus-one competitions are collected from internet. Three kinds of data are collected: (i)player

Table 1. Nature of Testing Data Set

Data Set	1	2	3	4	5
Player race	Undead	Undead	Orc	Orc	Elf
Enemy unit	Fm, P	Dr, Fm, P	Pr, So, Sb, P	Fm, P	Fm, P
No. of case	162	65	1004	549	869
No. of Combination	23	16	60	31	31

unit production statistics, (ii)enemy unit type, and (iii)performance scores. Five data sets with largest amount of cases are prepared as 1.

GA is a widely used technique for determining fuzzy measure due to its robustness. In this research, a comparison is conducted between our FELM and GA. In the GA algorithm, each chromosome consists $2^n - 1$ fuzzy measures with n unit types. The fitness function is designed to obtain the average differences between real scores and estimated scores. The detail setting of GA please refer to [2].

4.2 Comparison with Genetic Algorithm

All the simulations are performed in MATLAB. The machine used is an Intel Premium 4 2.3GHZ with 2GB Ram PC. 70% cases are used for training and the rest are for testing. The training time, training error, testing error are shown in Table 2. According to the results, we provide the following analysis.

- Obviously, the ELM algorithms are extremely fast, i.e., at most 1000 times faster than GA. It should be able to meet the requirement of real-time evaluating game strategies.
- In general, the training and testing errors of FELMs and GAs are similar. Moreover, we could see that GA gets a bad result with the data set 3(i.e. the biggest data set) but ELMs able to maintain the performance. Mathematically, if the valid training data are enough, the ELM algorithm is capable of learning a entirely fitted fuzzy measure.
- The Me-based FI could gets a higher accuracy than CI but time consuming by using GA . This is because it requires searching the whole power set of unit types in set selection process. The newly designed set selection algorithm of Me-based FI in ELM significantly reduces the complexity, made it with lower complexity and faster convergence than CI. Therefore, the combination of Me-based FI and ELM should be the best evaluator in four.

Table 2. Comparisons of FELM and GA

		Data set 1		Data set 2		Data set 3		Data set 4		Data set 5	
		Choquet	Mean	Choquet	Mean	Choquet	Mean	Choquet	Mean	Choquet	Mean
Training time	ELM	3.479s	0.390s	1.201s	0.172s	305.637s	37.222s	19.360s	2.340s	31.122s	3.666s
	GA	22.906s	1712s	10.644s	1572s	183.96s	34450s	64.723s	3727s	96.047s	4014s
Training error	ELM	0.115	0.144	0.084	0.190	0.042	0.046	0.121	0.137	0.053	0.054
	GA	0.194	0.138	0.201	0.181	0.485	0.078	0.195	0.137	0.059	0.061
Testing error	ELM	0.179	0.179	0.350	0.264	0.064	0.064	0.154	0.154	0.049	0.048
	GA	0.209	0.185	0.325	0.236	0.483	0.064	0.194	0.145	0.048	0.054

5 Conclusion

In this research, we propose a novel strategy evaluator in RTS game by combining Me-based FI and ELM. The experimental results demonstrate its feasibility and effectiveness. It appears to be a good choice of real-time evaluating different unit combination strategies in game. The future work may focus on developing an autonomous real-time systems capable of generating appropriate unit movement based on the current strategy evaluation model.

Acknowledgements. This research project is supported by the HK Polytechnic University grants 4-ZZAH.

References

1. Li, Y.J., Ng, H.F., Wang, H.B., Li, Y., Shiu, C.K.S.: Applying fuzzy integral for performance evaluation in real time strategy game. In: Proceedings of 2010 2nd International Conference on Information and Multimedia Technology (ICIMT 2010), Hong Kong, China, pp. 28–30 (December 2010)
2. Li, Y.J., Ng, H.F.P., Wang, H.B., Shiu, C.K.S., Li, Y.: Apply different fuzzy integrals in unit selection problem of real time strategy game. In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 170–177. IEEE (2011)
3. Aha, D.W., Molineaux, M., Ponsen, M.: Learning to win: Case-based plan selection in a real-time strategy game. In: Muñoz-Ávila, H., Ricci, F. (eds.) ICCBR 2005. LNCS (LNAI), vol. 3620, pp. 5–20. Springer, Heidelberg (2005)
4. Hsieh, J.L., Sun, C.T.: Building a player strategy model by analyzing replays of real-time strategy games. In: IEEE International Joint Conference on Neural Networks, IJCNN 2008, IEEE World Congress on Computational Intelligence, pp. 3106–3111. IEEE (2008)
5. Hagelbäck, J., Johansson, S.J.: A multi-agent potential field based bot for a full rts game scenario. Proceedings of Artificial Intelligence and Interactive Digital Entertainment, AIIDE (2009)
6. Sugeno, M.: Theory of fuzzy integrals and its applications (1974)
7. Wang, Z., Yang, R., Lee, K.H., Leung, K.S.: The choquet integral with respect to fuzzy-valued signed efficiency measures. In: 2008 IEEE International Conference on Fuzzy Systems, pp. 2143–2148. IEEE (2008)
8. Narukawa, Y.: Modeling decisions: information fusion and aggregation operators. Springer (2007)
9. Huang, G.B., Wang, D.H., Lan, Y.: Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics 2(2), 107–122 (2011)
10. Wang, J., Wang, Z.J.: Using neural networks to determine sugeno measures by statistics. Neural Networks 10(1), 183–195 (1997)

New Fuzzy Integral for the Unit Maneuver in RTS Game

Peter Hiu Fung Ng, YingJie Li, and Simon Chi Keung Shiu

Department of Computing, The Hong Kong Polytechnic University, Hong Kong
{cshfng,csyjli,csckshiu}@comp.polyu.edu.hk

Abstract. A new strategy planning is proposed to improve the effectiveness of unit maneuver in RTS game. Following several other researchers, our approach is based on the technique of potential field to provide an efficient path searching in dynamic environment. However, we adopt the Fuzzy Measure and Integral to indicate the interaction among the unit types. Hence, we also propose a new Fuzzy Integral to improve the quality of unit maneuver. Diversion and flank attack can be planned in potential field dynamically. We implement and present our techniques on real RTS game platform. The result is promising in this complex environment.

Keywords: Fuzzy Measure, Fuzzy Integral, Potential Field, CMA-ES, RTS.

1 Introduction

Unit formation and target of attack is the cores of unit maneuver in real time strategy (RTS) game. It is complicated as it consists of a great quantity of possibilities. Multiple targets and the intransitive superiority of unit formation lead the unit maneuver to remain a problem. Traditional tree searching or A* searching is unable to handle these two properties. There are too many weightings and each of them will interact with the others. In this research, we apply potential field, Fuzzy Measure and Integral to solve this problem. Potential field is suitable for complicated and various environment with multiple targets. However, it does not consider non-additive property. We integrate it with Fuzzy Measure and Integral to extend simple additive property to non-additive property. It provides the ability to handle interaction among different targets. We also propose a new Fuzzy Integral, Directional based Fuzzy Integral, to support the flank and diversion attack in unit formation planning.

2 Literature Review

More and more researchers are interested in RTS game and have recently generated many important outputs in the community [1, 2, 3]. They shared the same view on RTS game environment. It is hostile and dynamic. Counter maneuvers need to be implemented under uncertainty and time pressure. The classic potential field consists of two virtual forces [4, 5]. One is attractive potential force which is used to represent the goal. Another force is repulsive potential force which is used to represent the obstacle. Multi-agents potential fields of Hagelback [6] showed the importance to unit maneuver and tactics development.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 256–261, 2013.

Fuzzy Measure and Integral have been proven the ability of interaction handling [7, 8]. We have also implemented Fuzzy Integral to evaluate the combined power of units [9, 10, 11]. It is an efficient way to handle the interaction problem in RTS game. However, the aggregator of Fuzzy Integral has not been investigated and visualized very well. Our proposed idea here is to measure the opponent power by the Fuzzy Integral and provide a planning model for unit maneuver.

3 Methodology

The main ideas of this model are to minimum the interaction of enemy units and maximize the interaction of player units. First, the weighting of Fuzzy Measure for different unit types are learnt from real RTS game data. Then, the weightings are assigned to a potential field. All potentials of different unit types are summed up with geometric functions. Finally, the highest potential will become the next movement target.

3.1 Learning Fuzzy Measure from Real RTS Game Data

We use the Warcraft III replay to learn the interaction from real data [6]. For each replay r, unit proportions and scores of player and enemy are extracted. $X = \{x_1, x_2, \ldots, x_n\}$ is used to represent the enemy unit proportion, where x_i is the proportion of unit type i and n is the total number of unit type. $Y = \{y_1, y_2, \ldots, y_n\}$ is used to represent the player unit proportion. These quantities here are measured by the amount of resources used instead of the physical quantity count. Then it is normalized by a normalization function f .The scores, $Score_{player}(r)$ and $Score_{enemy}(r)$ are also extracted from each replay r .

Fuzzy Measure is used to represent the combined power of enemy and player. Fuzzy measure of enemy is defined as a mapping: $\mu: P(X) \rightarrow [0, 1]$, where $P(X)$ is the power set of X, i.e., all the (2^n-1) subsets of X. The Fuzzy Measure of player is defined as a mapping: $\mu: P(Y) \rightarrow [0, 1]$ with the same condition. $\mu(X)$ and $\mu(Y)$ are trained separately by CMA-ES and Fuzzy Integral. First, the $\mu(X)$ or $\mu(Y)$ are randomly assigned. Then the Fuzzy Integral, $Score_{player}(r)$ and $Score_{enemy}(r)$ are used to be the fitness function. The solution is optimized throughout the iteration. Detail procedures are shown in our previous researches [9, 10, 11].

3.2 Combining Choquet Integral and Potential Field

The contribution of each unit type in CI should be investigated and defined as a point of interest (POI) in the potential field. CI is converted into another form as shown in Equation (1). Assume a_i is the sorted x_i. $C(a_i)$ is used to represent the contribution of unit type a_i . Its equation is shown in Equation (2). As the Fuzzy Measure is non-monotonic, the contribution of unit type, $C(a_i)$, may be negative. For example, if $\mu(a_2,a_3) < \mu(a_3)$, then $C(a_2)$ will be negative. Therefore, the negative interaction of enemy could be identified.

$$\sum_{i=1}^{n} (a_i) \cdot [\mu(x \mid f(x) \geq a_i) - \mu(x \mid f(x) \geq a_{i+1})] \tag{1}$$

$$C(a_i) = (a_i) \times \left[\mu(x \mid f(x) \geq a_i) - \mu(x \mid f(x) \geq a_{i+1})\right] \tag{2}$$

3.3 New Integral for Potential Field

Although the Choquet Integral could present the positive and negative interaction among the unit type, the priory sorting of *f(x)* is difficult for the game developer to understand. Hence, the meaning of $C(a_i)$ which is related to the subset of a_i and its upper term a_{i-1} is also difficult to understand. The individual contribution of a_i and the interaction to the other units are mixed together. Therefore, we develop Directional based Fuzzy Integral (Directional based FI). It is used to describe the correlation of two unit types. The entire equation is shown in Equation (3). The integral is divided into two parts. The first part, *O*, is its individual contribution. The second part, *I*, is used to describe all the correlation of unit types.

$$(d)\int f(x)d\mu = \sum_{i=1}^{n}(O+I) \tag{3}$$

$$\text{where} \quad O = f(x_i)\mu(x_i)$$

$$I = w\sum_{j=1,j\neq i}^{n} f(x_i)[\mu(x_i,x_j)-\mu(x_j)]$$

$$f(x_i)\neq 0, f(x_j)\neq 0$$

A potential field with size $S \times S$ is generated. The equation of each point in potential field, *P(x,y)*, is shown in the Equation (4). Each unit type is a point of interest (POI). The unit type's coordinate, P_i, is extracted in the battle. By the Equation (3), the individual contribution and interaction can be indicated independently. Then, they are combined with the decay function, φ which is a form of Euclidean distance. *D* is a weighting to control area of the affected. The larger the *D*, the smaller affected area.

$$P(\mathrm{x},\mathrm{y}) = \begin{cases} \sum_{i=1}^{n}(O\times\varphi(P(\mathrm{x},\mathrm{y}),P_i)) + I\times\varphi(P_i,P_j)) & \text{(4.1), (4.2) is true} \\ \sum_{i=1}^{n}(O\times\varphi(P(\mathrm{x},\mathrm{y}),P_i)) & \text{Otherwise} \end{cases} \tag{4}$$

where n is the total number of unit type, P_i is the coordinate of a_i

$$\varphi(P(\mathrm{x},\mathrm{y}),P_i) = (D\times\log(\sqrt{[P(\mathrm{x})-P_i(\mathrm{x})]^2+[P(\mathrm{y})-P_i(\mathrm{y})]^2}))^{-1}$$

$$\varphi(P_i,P_j) = (D\times\log(\sqrt{[P_i(\mathrm{x})-P_j(\mathrm{x})]^2+[P_i(\mathrm{y})-P_j(\mathrm{y})]^2}))^{-1}$$

D is the weighting to control area of affected

$$\sqrt{[P(\mathrm{x})-P_j(\mathrm{x})]^2+[P(\mathrm{y})-P_j(\mathrm{y})]^2} \leq \sqrt{[P_i(\mathrm{x})-P_j(\mathrm{x})]^2+[P_i(\mathrm{y})-P_j(\mathrm{y})]^2} \tag{4.1}$$

$$\theta(P_i,P_j)-\alpha \leq \theta(P(x,y),P_i) \leq \theta(P_i,P_j)+\alpha \tag{4.2}$$

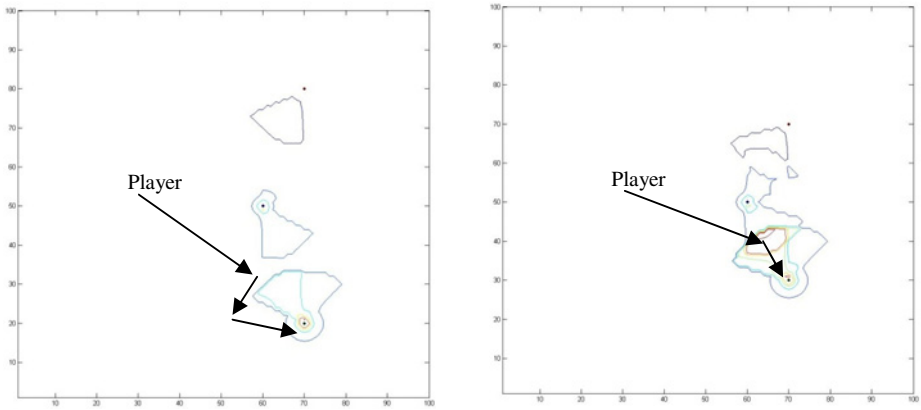
$$\text{where} \quad \theta(P_1,P_2) = \arctan\left(\frac{P_1(y)-P_2(y)}{P_1(x)-P_2(x)}\right) \quad \text{and} \quad \alpha = \pi/4$$

The affected area of individual contribution, O, is presented as a circle which is similar to the original potential field. However the affected area of interaction part, I, is presented as a sector of circle and points to another unit. The interaction part is generated in the area in between two points as the expression (4.1) and within a degree, α, as the expression (4.2).

The potential, I, represents the contribution of x_i to x_j. $\varphi(P_i,P_j)$ is a decay function which is used to enlarge the potential if the enemy unit is closer to each other. Figure 1(a) show the combined potential of two unit types. If P_1 and P_2 are getting away, the potential will drop sharply as shown in Figure 1(b).

(a) (b)

Fig. 1. Effect of the decay function, $\varphi(P_i,P_j)$



(a) (b)

Fig. 2. Effect of the decay function, $\varphi(P_i,P_j)$

3.4 Flanking and Diversion Attack

After the enemy and player potential field are generated, scripting is needed to decide the action of each unit type, such as flanking, diversion, direct attack or cooperation. First the highest potential of enemy and player are compared. One of them is selected. If the potential of player is higher, cooperation is performed to maintain the unit formation. On the other hand, if the potential of enemy is higher, the player will perform the flank (Figure 2(a)) or diversion attack (Figure 2(b)). The highest potential will

become the destination and the path is generated. All the process is updated at certain time slot. Destination and path is continually updated.

4 Experimental Result and Discussion

To prove the performance of potential field and fuzzy integral, we used Warcraft III to simulate the experiment. 1553 replays are collected from official competitions and used for training the Fuzzy Measure.

For CMA-ES, 50 search points are generated randomly. Their fitness is computed by Fuzzy Integral and compared with the scores of real game data. 25 better solutions are selected to computer mean vector, step size and covariance matrix. The search points are updated by these three parameters for the next generation. After 2500 iteration, the Fuzzy Measures will be used for testing.

A battle field is generated. Three unit types are randomly selected for each side. The number of unit in each type is randomly assigned but the total amount of the army is fixed to 20. Three unit maneuver models are tested. The first one is rule based system of Warcraft III. The second one is potential field and Choquet Integral. The last one is potential field and Directional based FI. For each experiment, 50 battles are preformed. The result has been stated in Table 1. The winning percentage of potential field with Directional based fuzzy integral is the highest. Compared with rule based system, potential field with Choquet Integral has a 29% improvement and the potential field with Directional based fuzzy integral has a 48% improvement.

The performance of unit maneuver is optimized. By observing the battle, potential fields with Directional based fuzzy integral showed a significant improvement when flank and diversion attack can be performed.

Table 1. Comparison of different Unit Maneuver

Unit Maneuver	Wining
Rule based System of Warcraft III	42%
Potential field and Choquet Integral	54%
Potential field and Directional based fuzzy Integral	62%

5 Conclusion

In this paper, we have developed an adversarial real time planning model for unit maneuver in RTS game. We have extended the normal additive properties to non-linear for potential field. We also defined a new type of fuzzy integral called Directional-based FI. The performance of path finding is optimized. Details unit maneuver, such as flank and diversion attack can be performed. The future work will focus on extending the result of feature interaction to perform more advance unit maneuver in RTS game, such as tracking problem.

References

1. Buro, M.: Call for AI research in RTS games. In: Proceedings of the AAAI 2004 Workshop on Challenges in Game AI, pp. 139–142 (2004)
2. Lucas, S.M.: Computational Intelligence and AI in Games. IEEE Transaction on Computational Intelligence and AI in Games 1, 1–3 (2009)
3. Alexander, N.: Game AI is Dead. Long Live Game AI! IEEE Intelligent Systems 22(1), 9–11 (2007)
4. Weijun, S., Rui, M., Chongchong, Y.: A Study on Soccer Robot Path Planning with Fuzzy Artificial Potential Field. In: 2010 International conference on Computing, Control and Industrial Engineering, pp. 386–390 (2010)
5. Yin, L., Yin, Y.: An Improved Potential Field Method for Mobile Robot Path Planning in Dynamic Environment. Proceeding of the 7th World Congress on Interlligent Control and Automation, 4847–4852 (2008)
6. Hagelback, J., Johansson, S.: Using multi-agent potential fields in real-time strategy games. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, vol. 2, pp. 631–638 (2008)
7. Ishii, K., Sugeno, M.: A model of human evaluation process using fuzzy measure. Int. J. Man-Machine Studies 22, 19–38 (1985)
8. Murofushi, T., Sugeno, M.: The Choquet integral in multiattribute decision making. Fuzzy Measures and Integrals: Theory and Applications, 333–347 (2000)
9. Li, Y.J.: Apply Different Fuzzy Integrals in Unit Selection Problem of Real Time Strategy Game. In: Proceeding of 2011 IEEE International Conference on Fuzzy System (Fuzz-IEEE 2011), Taipei, Taiwan, June 27-30, pp. 170–177 (2011)
10. Ng, P.H.F., Li, Y.J., Wang, H.B., Shiu, S.C.K.: Bottom-Up Strategy Planning Model by applying Fuzzy Integral in RTS Game. In: Proceeding of Joint 5th In-ternational Conference on Soft Computing and Intelligent Systems and 11th International Symposium on Advanced Intelligent Systems (SCIS & ISIS), Okayama, Japan, December 8-12, pp. 1579–1584 (2010); Best Student Paper Award
11. Li, Y.J., Ng, P.H.F., Wang, H.B., Li, Y., Shiu, S.C.K.: Applying Fuzzy Integral for Performance Evaluation in Real Time Strategy Game. In: Proceeding of 2010 2nd International Conference on Information and Multimedia Technology (ICIMT), Hong Kong, RPC China, December 28-30, pp. 168–172 (2010)

Weighted Coordinate-Wise Pegasos

Vilen Jumutc and Johan A.K. Suykens

Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001, Leuven, Belgium
{vilen.jumutc,johan.suykens}@esat.kuleuven.be

Abstract. Pegasos is a popular and reliable machine learning algorithm for making linear Support Vector Machines solvable at the larger scale. It benefits from the strongly convex optimization objective, faster convergence rates and lower computational and memory costs. In this paper we devise a new weighted formulation of the Pegasos algorithm which favors from the different coordinate-wise λ_i regularization parameters. Together with the proposed extension we give a brief theoretical justification of its convergence to an optimal solution and analyze at a glance its computational costs. We conclude our paper with the numerical results obtained for UCI datasets and demonstrate the merits and the importance of our approach for achieving a better classification accuracy and convergence rates in the partially or fully stochastic setting.

1 Introduction

Recent advances in linear Support Vector Machines (SVM) and the first order stochastic optimization [1–4] unlocked a room for improvement in the large-scale machine learning and black-box modelling. Usually computation of a full gradient for such large-scale problems is not bearable on a single machine, so one might consider to use a stochastic approximation to the original problem. In expectation [2, 4] such approach converges to the global optima if we consider strongly convex optimization objectives. Additionally the latter approach considerably saves the memory and only moderately increases the number of iterations to converge.

In this paper we propose a new weighted formulation of the Pegasos optimization objective and revise some of the algorithmic steps in order to maintain the consistency with the underlying theory. The key feature of the Pegasos [2] algorithm is a strongly convex optimization objective with the proper projection step. This combination helps to achieve a solution of accuracy ϵ in $O(\frac{R^2}{\lambda\epsilon})$ iterations where λ is the regularization parameter. On the other hand this approach imposes uniformed convergence speed for every single dimension. In general such uniform convergence is useful but doesn't reflect the importance and contribution of each dimension to the final classification result. The latter problem in Bayesian inference relates to Automatic Relevance Determination (ARD) [5–7]. We will compare this approach with the tuning via cross-validation for finding optimal λ_i hyperparameters.

The task of inferring optimal tuning parameters for our weighted coordinate-wise Pegasos is interesting by itself. In this paper we will highlight recent research

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 262–269, 2013.

in this subject and stress the difference in assumptions and behavior of Bayesian approach and tuning via global optimization techniques like Coupled Simulated Annealing (CSA) [8]. In our formulation of the Pegasos algorithm we deal with the linear approach and we are interested in obtaining the fittest regularization parameters λ_i *w.r.t* generalization and training speed-up.

This paper is organized as follows. Section 2 outlines original and Weighted Coordinate-Wise Pegasos formulations, explains in detail our algorithm and gives a theoretical background, *i.e.* convergence proofs. Section 3 highlights Automatic Relevance Determination and Coupled Simulated Annealing used for tuning λ_i regularization hyperparameters. Experimental setup and numerical results are given in Section 4 while Section 5 concludes the paper.

2 Weighted Coordinate-Wise Pegasos

Our approach is based on a simple assumption related to Automatic Relevance Determination. If we impose regularization differently for each dimension or data point we end up with the method which has a weighted impact of the features, dimensions or data samples on the classification output. In the case of the dual representation of SVMs we might talk about box constraints [9] which are related to the dual α unknowns. The larger is the upper bound of the particular box constraint the greater might be the influence of that data point on the decision boundary. On the other hand with ARD techniques and Bayesian inference one maximizes the posterior and the evidence can be used to assign a preference to alternative values of the hyperparameters [5].

To obtain such weighted formulation for the Pegasos algorithm first we propose a new optimization objective

$$f_{wcw}(w; \mathcal{A}_t) = \frac{1}{2} w^T \Lambda w + \frac{1}{|\mathcal{A}_t|} \sum_{(x,y) \in \mathcal{A}_t} \mathbb{L}(w; (x, y)), \tag{1}$$

where f_{wcw} stands for our new "weighted coordinate-wise" instantaneous optimization objective and it differs *w.r.t* original Pegasos objective only in the way we define our hyperparameters. Here Λ stands for the diagonal matrix with entries corresponding to coordinate-wise λ_i regularization hyperparameters.

Now we can see that the subgradient term becomes

$$\nabla_t = \Lambda w_t - \frac{1}{|\mathcal{A}_t|} \sum_{(x,y) \in \mathcal{A}_t^+} yx. \tag{2}$$

These modifications to the Pegasos algorithm do not guarantee a convergence without further modifications to the projection step defined in the original Pegasos as a projection onto the set $\mathcal{B} = \{w : \|w\| \leq 1/\sqrt{\lambda}\}$. In the next subsection we will take a closer look at our Weighted Coordinate-Wise Pegasos algorithm and present these missing details.

Algorithm 1. Weighted Coordinate-Wise Pagasos

Data: $\mathcal{S}, \Lambda, T, k, \epsilon$
1 Set $\lambda_{min} = \min_i \Lambda_{ii}$
2 Select w_1 randomly s.t. $\|w^{(1)}\| \leq 1/\sqrt{\lambda_{min}}$
3 **for** $t = 1 \rightarrow T$ **do**
4 Set $\eta_t = \frac{1}{t}\Lambda^{-1}$
5 Select $\mathcal{A}_t \subseteq \mathcal{S}$, where $|\mathcal{A}_t| = k$
6 $\rho = \frac{1}{|\mathcal{S}|}\sum_{(x,y)\in\mathcal{S}}(y - \langle w_t, x\rangle)$
7 $\mathcal{A}_t^+ = \{(x,y) \in \mathcal{A}_t : y(\langle w_t, x\rangle + \rho) < 1\}$
8 $w_{t+\frac{1}{2}} = w_t - \eta_t(\Lambda w_t - \frac{1}{k}\sum_{(x,y)\in\mathcal{A}_t^+} yx)$
9 $w_{t+1} = \min\left\{1, \frac{1/\sqrt{\lambda_{min}}}{\|w_{t+\frac{1}{2}}\|}\right\} w_{t+\frac{1}{2}}$
10 **if** $\|w_{t+1} - w_t\| \leq \epsilon$ **then**
11 **return** $(w_{t+1}, \frac{1}{|\mathcal{S}|}\sum_{(x,y)\in\mathcal{S}}(y - \langle w_t, x\rangle))$
12 **end**
13 **end**
14 **return** $(w_{T+1}, \frac{1}{|\mathcal{S}|}\sum_{(x,y)\in\mathcal{S}}(y - \langle w_t, x\rangle))$

2.1 The Algorithm

Hereby we present a brief summary of the modified Pegasos in Algorithm 1. Next we continue with the theoretical guarantees for convergence of this algorithm in the next subsection. In Algorithm 1 we can see a major "for" loop where gradient and projection steps are taking place and a minor "if" condition terminates an execution if the norm of the difference of two subsequent w vectors is less than ϵ. In Algorithm 1 we denote the whole dataset by $\mathcal{S}$ and at each iteration select randomly k samples for computation of the subgradient in Eq.(2). We should emphasize the importance of the projection step at the Line 9 where we are projecting our solution back on the ball with the radius $1/\sqrt{\lambda_{min}}$ and updated step size at the Line 4.

2.2 Analysis

This subsection presents a convergence analysis and theoretical background which brings our algorithm to the similar bounds as in the original paper of Shalev-Shwartz *et al.* [2]. We extend the analysis given in their paper to our new instantaneous objective by presenting Theorem 1.

Theorem 1. *Let $\mathcal{B}$ be a closed convex set and define $\prod_{\mathcal{B}}(w) = arg\min_{w'\in\mathcal{B}}\|w - w'\|$. Let $w_1, \ldots, w_T$ be a sequence of vectors such that $w_1 \in \mathcal{B}$ and for $t \geq 1$, $w_{t+1} = \prod_{\mathcal{B}}(w_t - \eta_t\nabla_t)$, where ∇_t is a subgradient of Eq.(1) at w_t and $\eta_t = \frac{1}{t}\Lambda^{-1}$. Let w^* be the solution of the optimization problem in Eq.(1), $\lambda_{min} = \min_i \Lambda_{ii}$, $\lambda_{max} = \max_i \Lambda_{ii}$, $\lambda_{min} > 0$ and let $G = \lambda_{max}/\sqrt{\lambda_{min}} + R$. Assume $\|x\| \leq R$ for all $(x,y) \in \mathcal{S}$. Then, for $T \geq 1$ we have*

$$\|w_T - w^*\| \leq 2G\frac{\lambda_{max}}{\lambda_{min}^2}(\ln(T) + 1) + \frac{2R}{\lambda_{min}}.$$

Proof. To prove our theorem we refer to Lemma 4 in [11] and Section 9.1.2 in [12]. To apply results from Boyd's book [12] first we need to show that our problem is strongly convex. It is easy to verify that the first term in Eq.(1) is λ_{min}-strongly convex with respect to w. Since f is a sum of λ_{min}-strongly convex function and another convex function (hinge loss), it is also strongly convex. From Section 9.1.2 in [12] we know that one can bound the distance to the minimizer w^* of the objective in Eq.(1) by

$$\|w_T - w^*\| \leq \frac{2}{\lambda_{min}} \|\nabla f(w_T)\|.$$

The simplistic bound on $\|\nabla f(w_T)\|$ is derived using the upper-bound on the matrix-vector product as follows

$$\|\nabla f(w_T)\| \leq \|\Lambda\| \|w_T\| + \frac{1}{|\mathcal{A}_T|} \sum_{(x,y) \in \mathcal{A}_T^+} \|yx\| \leq \lambda_{max} \|w_T\| + R.$$

Next using the fact that there exists t-independent bound on $\|\nabla f(w)\| \leq G$, $w_t = w_{t-1} - \eta_t \nabla_t$, $w_0 = 0$ and applying induction and the triangular inequality one can show that

$$\|w_T\| \leq \sum_{t=1}^{T} \|\eta_t \nabla f(w_t)\| \leq G \|\Lambda^{-1}\| \sum_{t=1}^{T} \frac{1}{t} \leq \frac{G}{\lambda_{min}} (\ln(T) + 1).$$

Next by assuming $\mathcal{B} = \{w : \|w\| \leq 1/\sqrt{\lambda_{min}}\}$ and the fact that $\|x\| \leq R$ we can derive t-independent bound on subgradient $\nabla f(w)$. The explicit form for the subgradient evaluated at intermediate solution w_t is given by Eq.(2). Using the triangular inequality, taking into account that Λ is a diagonal matrix one obtains

$$\|\nabla f(w)\| \leq \|\Lambda\| \|w\| + \tfrac{1}{|\mathcal{A}_t|} \textstyle\sum_i \|x_i\| \leq \lambda_{max} \|w\| + R \leq \lambda_{max}/\sqrt{\lambda_{min}} + R.$$

Finally we have to show that $w^* \in \mathcal{B}$ and that there exist a universal upper bound for $\|w\|$. As a starting point we take Lemma 4 in [11]. First we need to derive the dual form of our instantaneous optimization objective for any subset $\mathcal{A}_t$ at iteration t. We use the fact that there exists a vector $\alpha^* \in [0,1]^m$ which maximizes our dual objective and duality gap equals to zero. Assuming these conditions and $|\mathcal{A}_t| = k$ one gets

$$\frac{1}{2} w^{*T} \Lambda w^* + \frac{1}{k} \sum_{(x,y) \in \mathcal{A}_t} \mathbb{L}(w^*; (x,y)) = -\frac{1}{2} w^{*T} \Lambda w^* + \frac{1}{k} \|\alpha^*\|_1.$$

Rearranging the above, using the non-negativity of the hinge loss, the bound on $\|\alpha\|_1 \leq k$ and the fact that $w^{*T} \Lambda w^* \geq \lambda_{min} \|w^*\|^2$ we obtain our initial condition of set $\mathcal{B}$: $\|w\| \leq 1/\sqrt{\lambda_{min}}$. Now we can plug-in everything back to the inequality in Theorem 1 and complete the proof by putting all together. ■

3 Obtaining λ_i Hypermarameters

In this section we focus on the problem of estimating λ_i hyperparameters. This problem is interesting by itself and gained a lot of attention within the framework of Automatic Relevance Determination in neural networks and kernel methods [5–7]. In our experiments with Weighted Coordinate-Wise Pegasos algorithm we will use two distinct strategies for obtaining these hyperparameters.

The first one is related to cross-validation and global optimization techniques, like Coupled Simulated Annealing [8]. Following this approach we evaluate 10-fold cross-validation for each iteration in every standard simulated annealing process. In CSA one couples together acceptance probability functions of these processes in order to control general statistical measures that may have crucial influence on the performance of the optimization.

The other technique is purely Bayesian and relies on the recent research in this area [13]. The main idea of the latter approach is to estimate these hyper-parameters from the data by first marginalizing over the coefficients w and then performing what is commonly referred to as evidence maximization or type-II maximum likelihood [6, 14, 15]. Mathematically, in the context of Weighted Coordinate-Wise Pegasos this formulation is equivalent to minimizing

$$\mathcal{L}(\lambda) \triangleq -\log p(y;\lambda) \equiv \log|\Sigma_y| + y^T \Sigma_y^{-1} y, \tag{3}$$

where a flat hyperprior on λ_i hyperparameters is assumed and $\Sigma_y \triangleq \gamma I + \Phi \Lambda \Phi^T$. Here Φ denotes evaluated dataset and is given by the original mapping in the input space, *i.e.* $\phi(x) = x$, where $(x, y) \in \mathcal{S}$ as it is given in Theorem 1. In this setting we are following the proposed optimization via Iterative Re-Weighted Minimum l_1 in [13], where one is estimating the λ_i unknowns via alternating update rules which can be derived by decoupling $\mathcal{L}(\lambda)$ using upper bounding functions. This approach also relates to Sparse Bayesian Learning (SBL) that has been successful in a variety of applications [6].

4 Experiments

4.1 Setup

For the first part of our experiments we used a 2-step procedure for tuning the λ_i hyperparameters in Algorithm 1. This procedure consists of Coupled Simulated Annealing [8] initialized with 5 random sets of parameters for the first step and the simplex method [16] for the second step. After CSA converges to some local minima we select a tuple of λ_i hyperparameters which attains the lowest cross-validation error and start the simplex procedure to refine our selection. On every iteration step for CSA and simplex method we proceed with a 10-fold cross-validation.

For the second part of our experiments we used the ARD approach described in Section 3. For learning hyperparameters we used the procedure described in Section 2.1 of [13] and we defined $\gamma = 10^{-3}$ parameter related to our flat

hyperprior in Eq.(3) (we can bind this parametrization to the numerical stability of the underlying approach). Finally we defined $\epsilon = 10^{-5}$ stopping criterion which controls the norm of the difference for two subsequent solutions. Once we drop below this threshold iterative re-weighting procedure halts. In this part we experimented mostly with small-scale UCI datasets because of the computational burden of Σ_y matrix inversion and storage.

All experiments with large-scale UCI datasets were repeated 50 times with the random split to training and test sets in proportion 1:1. For the smaller UCI datasets we performed 50 iterations with splitting in proportion 1:9 taking 90% of the dataset for training. In the presence of 3 or more classes we performed binary classification where we learned to classify the first class versus all others. Only for Pen Digits dataset we performed several experiments where this setting was changed. For Algorithm 1 we fixed parameters: $T = 1000$ and $\epsilon = 10^{-5}$. Parameter k was set to $|\mathcal{S}|/10$ for all small-scale UCI datasets and we performed experiments with two different values of k for larger-scale UCI datasets.

4.2 Results

In this subsection we give numerical results on running Algorithm 1 within cross-validation and ARD settings. In Table 1 we can see our results for small-scale UCI datasets. Notation Pegasos$_{wcw}$ denotes our proposed approach with the tuning of λ_i hyperparameters via cross-validation and Pegasos$_{wcw/ard}$ stands for the same approach with hyperparameters obtained via the ARD method described in Section 3.

From Table 1 we can notice that Bayesian inference together with Algorithm 1 doesn't attain better results than the basic Pegasos algorithm. On the other hand Weighted Coordinate-Wise Pegasos performs better with respect to both of them. In Table 2a we can observe results for large-scale UCI datasets within

Table 1. Test errors for small-scale datasets

Dataset	Pegasos	Pegasos$_{wcw}$	Pegasos$_{wcw/ard}$
Ionosphere	0.16889	**0.12714**	0.28922
Parkinsons	0.20832	**0.18274**	0.23253
Sonar	**0.23848**	0.28667	0.32176
Iris	0.34933	**0.30133**	0.61467
Ecoli	0.06500	**0.05710**	0.16299

partially stochastic setting where we set k to be 10% of $|\mathcal{S}|$. We can notice that we perform equally better for almost all datasets in comparison with original Pegasos algorithm.

In Table 2b we can see that when going to a completely stochastic setting with $k = 1$ we are slightly deteriorating our performance. We can notice that our approach outperforms Pegasos algorithm for all datasets and it is well-suited for learning more accurately large scale linear SVMs via stochastic programming.

Table 2. Test errors for larger-scale datasets

(a) $k = 10\%$ of $|\mathcal{S}|$ (partially stochastic)

Dataset	Pegasos	Pegasos$_{wcw}$
Magic	0.30550	**0.23607**
Shuttle	0.21450	**0.08549**
Red Wine	**0.26924**	0.27747
White Wine	0.32443	**0.30541**
Covertype	0.36466	**0.32807**
Pen Digits (1 vs all)	0.10432	**0.08984**
Pen Digits (2 vs all)	0.08448	**0.05133**
Pen Digits (5 vs all)	0.09609	**0.06830**
Pen Digits (6 vs all)	0.05835	**0.02896**

(b) $k = 1$ (fully stochastic)

Dataset	Pegasos	Pegasos$_{wcw}$
Magic	0.32288	**0.27743**
Shuttle	0.21751	**0.08598**
Red Wine	0.29552	**0.29224**
White Wine	0.32807	**0.30599**
Covertype	0.36474	**0.34962**
Pen Digits (1 vs all)	0.10409	**0.09255**
Pen Digits (2 vs all)	0.08437	**0.05317**
Pen Digits (5 vs all)	0.09635	**0.07059**
Pen Digits (6 vs all)	0.05863	**0.02847**

5 Conclusions and Remarks

In this paper we proposed an extension of the Pegasos algorithm which is suitable for learning linear SVMs at the larger scale. It performs considerably better in terms of the generalization error. We presented two different approaches for learning λ_i regularization hyperparameters and showed that cross-validation together with CSA attains better results than a Bayesian approach. The latter results can be explained by the nature of ARD-related data-dependent prior distribution and inconsistencies between our assumptions and observed data. Finally we gave numerical results which demonstrated the merits of the proposed methods and verified the importance of the coordinate-wise parametrization for linear SVMs in terms of generalization error and training speed-up.

Acknowledgments. This work was supported in part by the scholarship of the Flemish Government; Research Council KUL: GOA/11/05 Ambiorics, GOA/10/09 MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), IOF-SCORES4CHEM, several PhD/postdoc & fellow grants; Flemish Government:FWO: PhD/postdoc grants, projects: G0226.06 (cooperative systems and optimization), G.0302.07 (SVM/Kernel), G.0320.08 (convex MPC), G.0558.08 (Robust MHE), G.0557.08 (Glycemia2), G.0588.09 (Brain-machine) research communities (WOG: ICCoS, ANMMM, MLDM); G.0377.09 (Mechatronics MPC), G.0377.12 (Structured models), IWT: PhD Grants, Eureka-Flite+, SBO LeCoPro, SBO Climaqs, SBO POM, O&O-Dsquare; Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011); IBBT; EU: ERNSI; ERC AdG A-DATADRIVE-B, FP7-HD-MPC (INFSO-ICT-223854), COST intelliCIS, FP7-EMBOCON (ICT-248940); Contract Research: AMINAL; Other: Helmholtz: viCERP, ACCM, Bauknecht, Hoerbiger. Johan Suykens is a professor at KU Leuven, Belgium.

References

1. Joachims, T.: Training linear SVMs in linear time. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2006, pp. 217–226. ACM, New York (2006)
2. Shalev-Shwartz, S., Singer, Y., Srebro, N.: Pegasos: Primal Estimated sub-GrAdient SOlver for SVM. In: Proceedings of the 24th International Conference on Machine Learning, ICML 2007, New York, NY, USA, pp. 807–814 (2007)
3. Chapelle, O.: Training a support vector machine in the primal. Neural Computation 19, 1155–1178 (2007)
4. Nemirovski, A., Juditsky, A., Lan, G., Shapiro, A.: Robust stochastic approximation approach to stochastic programming. SIAM J. on Optimization 19(4), 1574–1609 (2009)
5. MacKay, D.J.: Comparison of approximate methods for handling hyperparameters. Neural Computation 11, 1035–1068 (1999)
6. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. J. Mach. Learn. Res. 1, 211–244 (2001)
7. Chu, W., Keerthi, S.S., Ong, C.J.: Bayesian trigonometric support vector classifier. Neural Comput. 15(9), 2227–2254 (2003)
8. Xavier-De-Souza, S., Suykens, J.A.K., Vandewalle, J., Bollé, D.: Coupled simulated annealing. IEEE Trans. Sys. Man Cyber. Part B 40(2), 320–335 (2010)
9. Vapnik, V.: Statistical learning theory, 1st edn. Wiley (September 1998)
10. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: Least Squares Support Vector Machines. World Scientific, Singapore (2002)
11. Shalev-Shwartz, S., Singer, Y.: Logarithmic regret algorithms for strongly convex repeated games. Technical report, The Hebrew University (2007)
12. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, New York (2004)
13. Wipf, D., Nagarajan, S.: A New View of Automatic Relevance Determination. In: Platt, J.C., Koller, D., Singer, Y., Roweis, S. (eds.) Advances in Neural Information Processing Systems 20, pp. 1625–1632. MIT Press, Cambridge (2008)
14. MacKay, D.J.: Bayesian interpolation. Neural Computation 4, 415–447 (1991)
15. Neal, R.M.: Bayesian Learning for Neural Networks. Springer-Verlag New York, Inc., Secaucus (1996)
16. Nelder, J.A., Mead, R.: A simplex method for function minimization. Computer Journal 7, 308–313 (1965)

Recognition of Cross Profiles of Roadbed Based on Polygonal Representations

Andrey G. Bronevich[1,2], Alexander E. Lepskiy[1], Vladimir I. Umansky[3], and Dmitry A. Yakushev[3]

[1] National Research University "Higher School of Economics", Moscow, Russia
[2] JSC Research, Development and Planning Institute for Railway Information Technology, Automation and Telecommunication, Moscow, Russia
[3] JSC Intechgeotrans, Moscow, Russia

Abstract. The paper is devoted to the description of two approaches for recognizing railway roadbed profiles, obtained with the help of laser scanning. The first approach is based on the identification of similar parts of comparing profiles, presented by their polygonal representations. The second approach uses weighted functional metrics, where it is possible to process incomplete data and then to make a choice on the set of preferences.

1 Introduction

Nowadays there are several program systems that allow to monitor automatically railway roads. These systems give possibility to find defects linked with the track structure, gabarite dimensions, rail cross profiles, etc. The full information about it can be found in [1, 2, 8, 9]. The development of such systems is produced in two directions with the help of using new hardware or new software to achieve higher monitoring speed and accuracy.

In the paper we consider the problem of roadbed profiles recognition that is then used for detecting their defects, like non-normative breadth of ballast shoulder, non-normative breadth of roadbed shoulder, places of oversized angles of slope. The review of the possible approaches of detecting such defects can be found in [1]. We assume that polygonal representations of measured cross profiles of the roadbed are the input data for our methods. These polygonal representations can be evaluated statistically [4] by using clouds of points obtained by laser devices.

2 Data Description: Cross Profiles of Roadbed

The ideal cross profile of the roadbed can be represented as a polygon. Meanwhile, a real cross profile considerably differs from an ideal one, but it is possible to approximate its geometrical form by the sequence of straight segments. As we mention before, there is a problem of finding defects of the roadbed, and it

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 270–276, 2013.

can be solved by comparing the measured profile with the normative profile according to the design decision. In Russia there are special normative laws called "Constructional Norms and Rules", see [6] for example. There are some differences in solving such a problem linked with prior information in our disposal. The ideal situation is when the design decision is known. Then it is necessary to compare the etalon profile based on the decision design with the measured one to detect the set of possible defects. If the design decision is not known we need to recognize the measured profile using the set of all possible etalon profiles. In this case we have the set of etalon profiles that correspond to different types of roadbed that can be classified as ditch cuts, embankments, tunnels, etc. Examples of such etalon profiles are shown on Fig. 1.

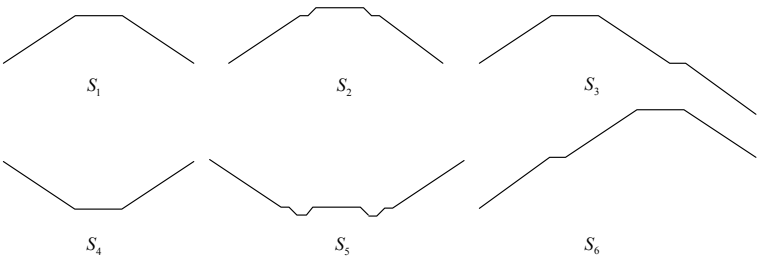


Fig. 1. Six etalon profiles

The analysis of the cross profile of the roadbed assumes that we need to extract feature points, whose positions determine the basic characteristics of the roadbed. The feature points are points corresponding to edges of upper and lower surfaces of embankment (ditch cut) of the roadbed and the ballast section. These points can be extracted by the statistical method of recovering profile described in [4]. It is worth to mention that such points are extracted with different plausibility and reliability. It was experimentally checked that points describing the ballast section are extracted with the better degree of plausibility.

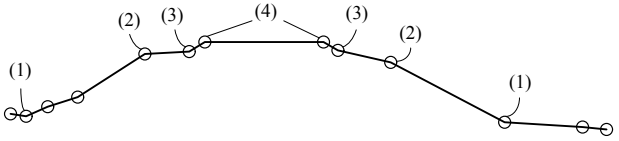


Fig. 2. The measured cross profile of the road bed with extracted feature points

The example of the measured profile is depicted on Fig. 2 with its polygonal representation. Each vertex of this representation can be interpreted as a feature

point. One can see there edges of lower (1) and upper (2) surfaces, edges of lower (3) and upper (4) surfaces of the ballast section.

3 Recognition of Road Bed Profiles Based on Their Polygonal Representations

3.1 The Problem Statement

Given etalon profiles $S_1, ..., S_N$ of the roadbed. Each profile S_j, $j = 1, ..., N$, is defined by a polygonal representation $S_j = \{\mathbf{s}_1^{(j)}, ..., \mathbf{s}_{n_j}^{(j)}\}$, where $\mathbf{s}_i^{(j)} = (u_i^{(j)}, v_i^{(j)})$, $i = 1, ..., n_j$, $j = 1, ..., N$, are vertices of the corresponding polygon. We assume that $u_1^{(j)} < u_2^{(j)} < ... < u_{n_j}^{(j)}$. It is necessary to recognize the measured profile that is also described by the polygonal representation $X = \{\mathbf{x}_1, ..., \mathbf{x}_m\}$, where $\mathbf{x}_i = (x_i, y_i)$, $i = 1, ..., m$, and $x_1 < x_2 < ... < x_m$.

The classification problem consists in finding etalon profile S_j that is similar to the measured profile X according to the chosen distant function. After that it is possible to compute the important characteristics of X like the height of embankment or the depth of ditch cut, and others.

We will describe below two algorithms for solution of this classification problem.

3.2 The Comparison of Polygonal Representations Based on Invariants

If we compare polygonal representations X and S_k, then they can differ in number of vertices, and in choice of coordinate system. To solve this problem, we propose to normalize them, i.e. to find a coordinate system, where considered polygonal representations are the most similar. This can be produced by the approach, described in [5]. Let $X^{(p,l)} = \{\mathbf{x}_i\}_{i=p}^{p+l-1}$ be a part of X that consists of l points. We will characterize any such representation $Y = X^{(p,l)}$ by the following vectors of features that are invariant w.r.t. the choice of coordinate system:

1) $\mathbf{r}(Y) = (r_i(Y))_{i=1}^{l-2}$, where $r_i(Y) = |\mathbf{r}_{i+2} - \mathbf{r}_{i+1}|/|\mathbf{r}_{i+1} - \mathbf{r}_i|$ is the ratio of lengths of two neighboring segments;

2) $\beta(Y) = (\beta_i(Y))_{i=1}^{l-2}$, where $\beta_i(Y)$ is the angle between neighboring segments $[\mathbf{r}_i, \mathbf{r}_{i+1}]$ and $[\mathbf{r}_{i+1}, \mathbf{r}_{i+2}]$;

3) $\sigma(Y) = (\sigma_i(Y))_{i=1}^{l-2}$, where $\sigma_i(Y)$ is the sign of angle between neighboring segments $[\mathbf{r}_i, \mathbf{r}_{i+1}]$ and $[\mathbf{r}_{i+1}, \mathbf{r}_{i+2}]$.

Let $X_1^{(p_1,l_1)}$ and $X_2^{(p_2,l_2)}$ be parts of polygonal representations. Then they are called *comparable* if $l_1 = l_2$ and $\sigma\left(X_1^{(p_1,l_1)}\right) = \sigma\left(X_2^{(p_2,l_2)}\right)$. Let $G_0 = \left\{\left(X^{(p,l)}, S_k^{(m,l)}\right)\right\}$ be the set of comparable parts of polygonal representations of the measured profile and etalon profiles. Then we can compute the similarity of comparable representations using the Euclidean metric

$$\rho\left(X^{(p_1,l)}, S_k^{(p_2,l)}\right) = w_1\sqrt{\left\|\mathbf{r}\left(X^{(p_1,l)}\right) - \mathbf{r}\left(S_k^{(p_2,l)}\right)\right\|^2 + w_2\left\|\beta\left(X^{(p_1,l)}\right) - \beta\left(S_k^{(p_2,l)}\right)\right\|^2},$$

where $w_1 > 0$ is the coefficient allowing us to increase the preference of some pairs (for example, pairs with a large number of segments), $w_2 > 0$ is the coefficient of scaling of vectors $\mathbf{r}$ and β.

Let

$$\Upsilon_k = \left\{ \rho\left(X^{(p_1,l)}, S_k^{(p_2,l)}\right) : \left(X^{(p_1,l)}, S_k^{(p_2,l)}\right) \in G_0 \right\}, k \in \{1, ..., N\},$$

be the ordered set of distances, where $\rho\left(X^{(p_1{}^{(1)},l^{(1)})}, S_k^{(p_2{}^{(1)},l^{(1)})}\right)$ is after $\rho\left(X^{(p_1{}^{(2)},l^{(2)})}, S_k^{(p_2{}^{(2)},l^{(2)})}\right)$ if $p_1^{(1)} > p_1^{(2)}$ or $\left(p_1^{(1)} = p_1^{(2)}\right) \wedge \left(l^{(1)} \geq l^{(2)}\right)$. According to the algorithm, we should detect local minima in vectors $\Upsilon_k = \{\rho_i\}$, $k = 1, ..., N$. The set of pairs, that correspond to these minima, is denoted by G_1. For each pair $\left(X^{(p,l)}, S_k^{(m,l)}\right) \in G_1$ we normalize polygonal representations X and S_k by finding the coordinate system that corresponds to similar parts $X^{(p,l)}$ and $S_k^{(m,l)}$. This leads to the following optimization problem

$$\left\| AX^{(p,l)} - S_k^{(m,l)} - B \right\|^2 \to \min$$

w.r.t. orthogonal transformation A and vector $B \in R^2$. The analytical solution of this optimization problem can be found in [5]. After that we can compute the normalized representation of X by the formula

$$\tilde{X} = AX^{(p,l)} - B.$$

After that we should find the best normalization among possible ones. This is produced by finding a distance between the normalized profile $\tilde{X}$ and the etalon profile S_k. One of possible ways to do this is based on distance images [7]. In this case polygons $\tilde{X}$ and S_k are depicted as binary images. Let $Y = \{y_{i,j}\}_{N_1 \times N_2}$, where $y_{i,j} \in \{0, 1\}$, be a binary image. Then the distance image is the matrix $Y^* = \{y^*_{i,j}\}_{N_1 \times N_2}$ computed by

$$y^*_{i,j} = \min_{(k,m)|y_{k,m}=1} \sqrt{(i-k)^2 + (j-m)^2}.$$

Let Y_1 and Y_2 be polygons and let ${Y_1}^*$ and ${Y_2}^*$ be the corresponding distance images. Then the closure between Y_1 and Y_2 can be measured by

$$\mu(Y_1, Y_2) = \frac{\|Y_1^* - Y_2^*\|}{\max\{\|Y_1^*\|; \|Y_2^*\|\}},$$

where $\|\ \|$ is the matrix norm. Let us notice that there are computationally effective algorithms that approximately compute distance images [3]. Let $\tilde{X}_k$ be the best normalization of X w.r.t. etalon image S_k providing the minimum of μ. Then the classification rule can be defined as follows: the measured profile X corresponds to the etalon S_m if $\mu(\tilde{X}_m, S_m) \leq \mu(\tilde{X}_k, S_k)$, $k = 1, ..., N$.

Let us assume that the set of all etalon profiles is depicted on Fig. 1, and the measured contour X is depicted on Fig. 2. Then applying the described

classification algorithm we get that the measured contour corresponds to the etalon S_2. Fig. 3 shows the similar parts of comparing polygonal representations and how the polygon $\tilde{X}_2$ is fitted to S_2.

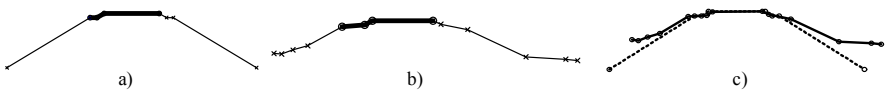


Fig. 3. Parts of polygonal representations: a) optimal $S_2^{(m,l)}$ (bold line); b) optimal $X^{(p,l)}$ (bold line); c) S_2 and $\tilde{X}_2$

3.3 The Comparison of Polygonal Representations by Norms on the Curve Space

Obviously, any polygon can be considered as a piecewise linear function. Therefore, there is a possibility for comparing polygons using a norm defined in a space of functions. For example, we can consider the normalized space $C_p^w[a,b]$ of continuous functions on $[a,b]$ with the norm $\|f\|_{w,p} = \left(\int_a^b w(t)|f(t)|^p dt\right)^{\frac{1}{p}}$, $1 \le p < \infty$, $\|f\|_{w,\infty} = \max\limits_{t\in[a,b]} w(t)\,|f(t)|$, where w is a non-negative weight function such that $\int_a^b w(t)dt = 1$. Let us assume that we can detect feature points $\mathbf{x}^{(0)} = (x^{(0)}, y^{(0)})$, $\mathbf{x}^{(1)} = (x^{(1)}, y^{(1)})$ of the measured profile X that correspond to edges of the upper surface of embankment or the lower surface of ditch cut. Obviously we can detect the corresponding points $\mathbf{s}^{(0)} = (u^{(0)}, v^{(0)})$ and $\mathbf{s}^{(1)} = (u^{(1)}, v^{(1)})$ in the etalon representations. Using these points we can produce the normalizations $\hat{S} = \{\hat{\mathbf{s}}_1, ..., \hat{\mathbf{s}}_n\}$ and $\hat{X} = \{\hat{\mathbf{x}}_1, ..., \hat{\mathbf{x}}_m\}$ of the etalon profile and the measured profile as follows:

$$\hat{u}_j = u_j - \tfrac{1}{2}(u^{(0)} + u^{(1)}), \hat{v}_j = v_j - \tfrac{1}{2}(v^{(0)} + v^{(1)}), \;\; j = 1, ..., n,$$

$$\hat{x}_i = \frac{\|\mathbf{s}^{(0)} - \mathbf{s}^{(1)}\|}{\|\mathbf{x}^{(0)} - \mathbf{x}^{(1)}\|}\left(x_i - \tfrac{1}{2}(x^{(0)} + x^{(1)})\right), \hat{y}_i = \frac{\|\mathbf{s}^{(0)} - \mathbf{s}^{(1)}\|}{\|\mathbf{x}^{(0)} - \mathbf{x}^{(1)}\|}\left(y_i - \tfrac{1}{2}(y^{(0)} + y^{(1)})\right),$$

$i = 1, ..., m$, where $\hat{\mathbf{s}}_j = (\hat{u}_j, \hat{v}_j)$, $j = 1, ..., n$; $\hat{\mathbf{x}}_i = (\hat{x}_i, \hat{y}_i)$, $i = 1, ..., m$.

For notation simplicity, let us assume that polygonal representations X and S are normalized. Then the distance between X and S can be computed by the following algorithm:

1) compute the union of sequences $\{x_i\}_{i=1}^m$ and $\{u_j\}_{j=1}^n$ such that $Z = \{z_k\}_{k=1}^l = \{x_i\}_{i=1}^m \cup \{u_j\}_{j=1}^n$: $z_1 < z_2 < ... < z_l$;

2) since any polygon is a piecewise linear function, the values of such functions are computed for profiles X and S for any point $z_k \in Z$. As result we get a profiles $X = \{\mathbf{x}_1, ..., \mathbf{x}_l\}$ and $S = \{\mathbf{s}_1, ..., \mathbf{s}_l\}$ correspondently, where $\mathbf{x}_i = (z_i, \tilde{y}_i)$, $\mathbf{s}_i = (z_i, \tilde{v}_i)$.

3) the transformations made in 1) and 2) allow us to compute the distances using formulas

$$d_{p,w}(X,S) = \sqrt[p]{\sum_{i=1}^{l-1} w_i |\tilde{y}_i - \tilde{v}_i|^p (z_{i+1} - z_i)}, \quad 1 \le p < \infty,$$

$$d_{\infty,w}(X,S) = \max_{1 \le i \le l} w_i \left|\tilde{y}_i - \tilde{v}_i\right|,$$

where weights $w_i > 0$ obeying $\sum_{i=1}^{l} w_i = 1$. Let us notice that the choice of w_i can decrease the influence of points with a small informativity that are located far from the track. If $w_i = 1$ for any i, then the corresponding metric is denoted d_p. Let us notice that the metric d_1 is proportional to quantity of work which must be made to correct of roadbed.

Let we apply metrics $d_1(X,S_k)$ and $d_{1,w}(X,S_k)$ with $w_i = w_0\left(1+\left|\frac{1}{2}l-i\right|^4\right)^{-1}$, $k = 1,...,6$, to the measured profile X on Fig. 2, and etalon profiles on Fig. 1. In this case the corresponding distances are given in Table 1.

Table 1. Values of distances $d_1(X,S_k)$ and $d_{1,w}(X,S_k)$ $k = 1,...,6$

	S_1	S_2	S_3	S_4	S_5	S_6
$d_1(X,S_k)$	41,5	28,9	78,8	105,9	77,5	157
$d_{1,w}(X,S_k)$	1,28	0,38	0,59	4,63	1,28	1,18

Thus, the etalon profile S_2 by both metrics has better fitted for X.

Next, we should select the most preferred profile. We say that the etalon profile S_i is more preferable than the profile S_j for the classification of the measured profile X if $d(X,S_i) < d(X,S_j)$. This fact is denoted by $S_i \succ S_j$. If we have $d(X,S_i) = d(X,S_j)$ then we say that these profiles have the same preferences and denote it by $S_i \simeq S_j$. It is possible to use the threshold comparison to enhance the robustness of the comparison procedure: $S_i \succ S_j$ if $d(X,S_i) \le d(X,S_j) - \varepsilon$ and $S_i \simeq S_j$ if $|d(X,S_i) - d(X,S_j)| < \varepsilon$, where $\varepsilon > 0$ is a threshold value.

Then we have the following ordering of etalon profiles (and related classes) by preference in according to the metric d_1: $S_2 \succ S_1 \succ S_5 \succ S_3 \succ S_4 \succ S_6$. We have the following ordering at 5% threshold comparison (i.e. ε is equal 5% from the $\max\limits_k d_1(X,S_k) - \min\limits_k d_1(X,S_k)$): $S_2 \succ S_1 \succ S_5 \simeq S_3 \succ S_4 \succ S_6$. The ordering of etalon profiles is different for the specified weighted metrics $d_{1,w}$: $S_2 \succ S_3 \succ S_6 \succ S_1 \simeq S_5 \succ S_4$ for the nonthreshold comparison and $S_2 \simeq S_3 \succ S_6 \simeq S_1 \simeq S_5 \succ S_4$ for 5% threshold comparison.

4 Summary and Conclusion

In the paper we present two algorithms for recognizing the cross profile of the roadbed. Clearly both algorithms have the similar procedures: 1) normalization

of polygonal representations; 2) the recognition of the measured profile based on distance function. The choice of the algorithm can be done taking into account the prior information (available or not information about feature points or other features that allows us to apply the simple algorithm for normalization) and stability of considered metrics to the noise.

Acknowledgment. The study was implemented in the framework of The Basic Research Program of the Higher School of Economics. This work was supported by the grants 11-07-00591-a and 1-07-13125-ofi-m-2011-RZhD of RFBR (Russian Foundation for Basic Research).

References

[1] Babenko, P.: Visual inspection of railroad tracks. In PhD Thesis (2009), http://server.cs.ucf.edu/vision/papers/theses/Babenko_Pavel.pdf
[2] Berry, A., Nejikovsky, B., Gilbert, X., Jajaddini, A.: High speed video inspection of joint bars using advanced image collection and processing techniques. In: Proc. of World Congress on Railway Research (2008)
[3] Borgefors, G.: Distance transformations in digital images. Computer Vision, Graphics, and Image Processing 34, 344–371 (1986)
[4] Bronevich, A.G., Karkishchenko, A.N., Umanskiy, V.I.: Statistical method of restoring the profile from laser scanning data. Digital Signal Processing 4, 42–49 (2011) (in Russian)
[5] Bronevich, A.G., Lepskiy, A.E.: Some effective approaches to the recognition of three-dimensional images of objects by the external contour. In: Proc. of the 7th National Conf. on Artificial Intelligence, Pereslavl-Zaleski, vol. 2, pp. 557–565 (2000) (in Russian)
[6] Constructional Norms and Rules 32-01-95. The railways of 1520 mm (1995) (in Russian)
[7] Liu, H.C., Srinath, M.D.: Partial Shape Classification Using Contour Matching in Distance Transformation. IEEE Trans. on Pattern Anal. and Mach. Intel. 11(11), 1072–1079 (1990)
[8] Machine Vision Inspection of Railroad Track. Technical report, NEXTRANS Project No.0281Y02 (2011)
[9] Trinh, H., Haas, N., Li, Y., Otto, C., Pankanti, S.: Enhanced Rail Component Detection and Consolidation for Rail Track Inspection. In: Proc. IEEE Workshop on the Applications of Computer Vision, pp. 289–295 (2012)

A New Orthogonalization of Locality Preserving Projection and Applications

Gitam Shikkenawis[1,⋆], Suman K. Mitra, and Ajit Rajwade[2]

[1] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat, India
[2] Indian Institute of Technology Bombay, Mumbai, Maharashtra, India
{201221004,suman_mitra}@daiict.ac.in, ajitvr@cse.iitb.ac.in

Abstract. Locality Preserving Projection (LPP) is a linear projection method that preserves the local structure to find the underlying manifold of the data. Non-orthogonality of LPP basis makes its use difficult not only for reconstruction but also for other applications such as denoising. At present, orthogonal basis of LPP (OLPP) are obtained in an iterative manner which is computationally expensive. In this article, a new orthogonalization of LPP (NOLPP) basis is proposed by relaxing the constraint used to minimize the objective function giving rise to the basis. The reducibility capacity of NOLPP for data clustering is validated by performing experiments on several databases. Use of NOLPP for image denoising shows its efficiency in comparison to the state of the art research. Fine structures present in the images are preserved even at high noise levels.

Keywords: Locality Preserving Projection, Dimensionality Reduction, Image Denoising.

1 Introduction

Data analysis and representation in a transformed domain has fetched a lot of attention. Many linear and non-linear methods have been proposed that try to learn the manifold structure from the given set of observations. Principle components analysis (PCA) [11], Linear Discriminant Analysis (LDA) [3] and Independent Component Analysis (ICA) [10] are the most widely used linear methods that learn the basis preserving global structure of the data. Locality Preserving Projection (LPP)[9], added in last decade, tries to preserve the local structure using the neighbourhood information. LPP is a linear method, but due to the characteristics of preserving the local structure, it captures the non-linearity present in the data. Variants of LPP [12], [6], [7] have also been proposed in recent years. Though, conventional LPP and its variants well preserve the local manifold structure, the basis vectors learnt are not orthogonal. Having orthonormal basis is advantageous in many applications and it makes the data

⋆ The author is acknowledging Tata Consultancy Services for providing the financial support for carrying out this work.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 277–283, 2013.

reconstruction much simpler. Orthogonal LPP [4] orthogonalized the basis of LPP in an iterative manner by selecting a basis vector orthogonal to previously learnt basis vectors. In this article, we are proposing a new orthogonalization of LPP (NOLPP) which learns orthogonal basis solving the eigen value problem only once.

LPP has been applied in various applications of dimensionality reduction, data classification and clustering, recognition tasks as well as document representation and indexing. Experiments of data clustering using the basis of currently proposed NOLPP have been performed. A new application of NOLPP in the area of Image Denoising has also been explored. Patches from the noisy image are used to learn the NOLPP basis and the denoising is performed in the transformed domain. As mentioned earlier, inverse transformation of the denoised patches into the spatial domain becomes much simpler because of the orthogonal property.

A brief discussion about conventional LPP, OLPP and the proposed NOLPP is included in Section 2. Section 3 presents clustering results on various datasets in using much less dimensions in the NOLPP domain. A new application of LPP for image denoising is covered in Section 4 along with the concluding remarks in Section 5.

2 Locality Preserving Projection

The objective function of LPP ensures closeness of the neighbouring data points in the projection domain by assigning them higher weights in the weight matrix $\mathbf{W}$. $\mathbf{x_i}$'s represent the data points in vector format and $\mathbf{y_i} = \mathbf{a^T x_i}$ are their projections on the learned basis $\mathbf{a}$ as follows:
$min \sum_{ij} \|\mathbf{y_i} - \mathbf{y_j}\|^2 \mathbf{W_{ij}} = min \sum_{ij} \left(\mathbf{a^T x_i} - \mathbf{a^T x_j}\right)^2 \mathbf{W_{ij}} = min\ \mathbf{a^T X L X^T a}$
Here, $\mathbf{X}$ is the matrix containing the data points ($\mathbf{x_i}'s$) in columns, $\mathbf{D} = \sum_i \mathbf{W_{ij}}$ is a diagonal matrix. $\mathbf{D} - \mathbf{W}$ is called the laplacian matrix i.e. $\mathbf{L}$. The weighing matrix $\mathbf{W}$ is computed as : $\mathbf{W_{ij}} \equiv e^{-\frac{\left\|\mathbf{x_i} - \mathbf{x_j}\right\|^2}{t}}$, if $\mathbf{x_i}$ belongs to k-NN of $\mathbf{x_j}$. Here, t controls the width of the Gaussian kernel. A normalization constraint $\mathbf{a^T X D X^T a = 1}$ is imposed on the given objective function which results in solving the generalized eigenvalue solution of the $\mathbf{X L X^T a} = \lambda \mathbf{X D X^T a}$. Eigenvectors corresponding to the minimum eigenvalues are the strongest basis vectors for the given dataset. Basis vectors learnt using LPP are not orthonormal. There are many applications of the projection based approaches where orthonormal basis vectors are more advantageous. Also, reconstruction of the data becomes difficult using non-orthonormal basis.

Orthonormal Locality Preserving Projection (OPLL) [4] imposed the orthonormality constraint on the basis vectors in the conventional LPP framework. The objective function now becomes, min $\mathbf{a^T X L X^T a}$ subject to the constraints $\mathbf{a^T X D X^T a = 1}$ and $\mathbf{a_m^T a_k} = 0, \forall m = \{1, 2, ..., k-1\}$.

The basis vectors $\mathbf{a_k}$s are found in the iterative manner assuring that the vector learnt is orthonormal to all the previously found basis vectors. $\mathbf{a_1}$ is the eigenvector associated with the smallest non-zero eigenvalue of $\mathbf{M^{(1)}} = \left(\mathbf{X D X^T}\right)^{-1} \mathbf{X L X^T}$. $\mathbf{a_k}, \forall k$ are computed as follows:

$$\mathbf{M}^{(k)} = \left\{ \mathbf{I} - \left(\mathbf{XDX^T}\right)^{-1} \mathbf{A}^{(k-1)} \cdot \left[\mathbf{B}^{(k-1)}\right]^{-1} \left[\mathbf{A}^{(k-1)}\right]^T \right\} \mathbf{M^{(1)}} \quad (1)$$

$\mathbf{a_k}$ is the eigenvector associated with the smallest eigenvalue of $\mathbf{M^{(k)}}$.

$$\mathbf{A}^{(k-1)} = [\mathbf{a_1}, \ldots, \mathbf{a_{k-1}}];\ \mathbf{B}^{(k-1)} = \left[\mathbf{A}^{(k-1)}\right]^T \left(\mathbf{XDX^T}\right)^{-1} \mathbf{A}^{(k-1)} \quad (2)$$

The basis matrix $\mathbf{a}$ thus found is orthonormal. Though, we get the orthonormal basis vectors, the method described above is computationally very expensive.

2.1 A New Orthogonalization of LPP

The fundamental need for the basis vectors to be orthogonal is $\mathbf{a^T a} = I$. Hence, in order to have orthogonal basis vectors, the constraint $\mathbf{a^T a} = I$ should be satisfied. We introduce this constraint while minimizing the objective function of conventional LPP as opposed to $\mathbf{a^T XDX^T a} = 1$. The minimization problem now reduces to the eigenvalue solution of $\mathbf{XLX^T a} = \lambda \mathbf{a}$. The solution is no more the generalized eigenvalue problem and produces orthonormal basis vectors directly in the first step, hence is computationally efficient. The computation of the weight matrix is carried out as per Extended LPP [12]. The weights are assigned using a z-shaped weighing function 3 that weighs the data points a monotonically decreasing fashion according to the Euclidean distance between them assuring their closeness in the projection domain as well as adding more discriminating power than the conventional LPP [12]. l and u are the lower and upper limits of the distance range within which the function changes its values.

The weighing function specified in Eqn. 3 is used throughout the article for experiments. l and u are the lower and upper limits of the distance range within which the function changes its values otherwise 1 and 0 values are directly assigned.

$$\mathbf{W}_{ij} \equiv \left\{ \begin{array}{ll} 1 - 2\left(\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2 - l}{u - l}\right)^2; & \text{if } l \leq \|\mathbf{x_i} - \mathbf{x_j}\|^2 \leq \frac{l+u}{2} \\ 2\left(\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2 - u}{l - u}\right)^2; & \text{if } \frac{l+u}{2} \leq \|\mathbf{x_i} - \mathbf{x_j}\|^2 \leq u \end{array} \right\} \quad (3)$$

3 Data Dimensionality Reduction and Clustering

Dimensionality Reduction is the key application of LPP and other such projection techniques. Several experiments have been performed on various datasets to test the data discrimination capacity of the proposed approach with very few dimensions. The results with varying no. dimensions (maximum 100) are included in order to show the capability of discrimination of the proposed approach with very less dimensions. Current approach is compared with the conventional LPP and one of its variants.

In case of data discrimination and clustering, ambiguities arise in the structurally similar data points. Sets of structurally similar data points from various

Table 1. Results (Errors) of clustering on various databases using LPP, ELPP and proposed NOLPP

Data	# Dimensions Used											
	2			10			50			100		
	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP	LPP	ELPP	NO-LPP
Digits 3-8	42.2	**17.0**	18.75	52.0	**14.5**	**14.5**	45.7	10.5	**4**	51.2	9.75	**3.25**
Digits 1-7	61.2	10.2	**3.25**	13.5	9.25	**3.25**	5.5	3.0	**1**	4.25	2.75	**1**
YALE Per 1,5	8.33	11.11	**6.94**	6.94	5.55	**2.78**	**0**	**0**	**0**	**0**	**0**	**0**
YALE Per 3,4	26.38	11.11	**8.33**	16.66	5.55	**2.77**	**1.38**	2.77	2.77	**1.38**	2.77	2.77
Person 1	72.47	3.89	**0.61**	71.98	0.85	**0**	70.52	**0**	**0**	67.11	**0**	**0**
Person 2	71.80	0.73	**0**	73.92	**0**	**0**	69.24	**0**	**0**	70.06	**0**	**0**

databases are selected for the experiments. The data points are projected on the learnt basis using LPP, ELPP and the current proposal and clustered using nearest neighbour approach. Digit pairs $3-8$ and $1-7$ from the MNIST database of Handwritten Digits [1] are chosen as they share similar structures. Two pairs of persons from the YALE Face DatabaseB [8] having resemblance are selected. Also, it is difficult to discriminate various expressions of a person, hence results of clustering four different facial expressions of two persons from the Video database [12] are shown in Table 1.

4 Image Denoising : A New Application

A new application of LPP has been explored for Image Denoising tasks. The basic noisy image formation model is $\mathbf{I}_\eta = \mathbf{I} + \eta$ where $\mathbf{I}$ is the original image, η is i.i.d. Gaussian noise with standard deviation σ and $\mathbf{I}_\eta$ is the noisy image. The procedure of removing noise is carried out in the projection domain of new Orthogonal LPP. Image is processed in the moving window fashion, considering a patch of size $n \times n$ at a time.

The noisy image formation model in the transformed domain using basis $\mathbf{a}$ can be given as, $\mathbf{I}_{\eta\mathbf{a}} = \mathbf{I}_\mathbf{a} + \eta_\mathbf{a}$. In order to remove noise, the coefficients of the noisy image in transformed domain $\mathbf{I}_{\eta\mathbf{a}}$ are modified using shrinkage rule such as Soft thresholding, Hard thresholding and Weiner filter update. Weiner filter update, $\frac{\sigma^2_{I_{\eta a}} - \sigma^2}{\sigma^2_{I_{\eta a}}}\mathbf{I}_{\eta\mathbf{a}}$ is used for all the experiments in this article as it produced the best results. Here, σ is the standard deviation of noise and $\sigma_{I_{\eta a}}$ is computed from the degraded image.

4.1 Experimental Results

The experiments are performed on greyscale texture and natural image databases. For all the experiments, the images are divided into overlapping patches of size 6×6. Each patch is considered as a data point for learning the NOLPP basis. For computation of weight matrix as per [12], the distance

Table 2. Denoising results - PSNR (SSIM): UIUC Texture Database ($\sigma \in \{20, 30, 40\}$)

		LPG-PCA1	LPG-PCA2	BM3D1	BM3D2	NOLPP
Bark	$\sigma = 20$	25.66 (0.79)	25.53 (0.78)	25.55 (0.78)	25.79 (0.79)	**26.04** (**0.82**)
	$\sigma = 30$	23.71 (0.68)	23.58 (0.66)	23.70 (0.67)	23.88 (0.68)	**24.26** (**0.74**)
	$\sigma = 40$	22.56 (0.61)	22.46 (0.58)	22.25 (0.58)	22.73 (0.61)	**23.08** (**0.67**)
Water	$\sigma = 20$	28.23 (0.76)	28.20 (0.76)	28.24 (0.76)	28.56 (0.78)	**28.96** (**0.80**)
	$\sigma = 30$	26.24 (0.65)	26.32 (0.66)	26.46 (0.66)	26.87 (0.69)	**27.17** (**0.72**)
	$\sigma = 40$	24.91 (0.56)	25.18 (0.58)	25.19 (0.58)	25.76 (0.63)	**25.91** (**0.65**)
Pebbles	$\sigma = 20$	26.81 (0.79)	26.75 (0.79)	26.70 (0.78)	27.05 (0.80)	**28.14** (**0.85**)
	$\sigma = 30$	24.74 (0.69)	24.70 (0.68)	24.82 (0.68)	25.15 (0.71)	**26.16** (**0.78**)
	$\sigma = 40$	23.44 (0.60)	23.46 (0.60)	23.35 (0.59)	23.92 (0.63)	**24.87** (**0.71**)
Brick	$\sigma = 20$	26.16 (0.80)	26.09 (0.79)	26.04 (0.79)	**26.28** (0.80)	26.02 (**0.82**)
	$\sigma = 30$	24.14 (0.70)	24.08 (0.69)	24.21 (0.70)	**24.43** (0.71)	24.17 (**0.72**)
	$\sigma = 40$	22.88 (0.63)	22.85 (0.61)	22.69 (0.61)	**23.23** (0.63)	23.00 (**0.65**)
Fur	$\sigma = 20$	28.36 (0.80)	28.41 (0.80)	30.63 (0.82)	**31.35** (**0.86**)	31.24 (0.85)
	$\sigma = 30$	26.18 (0.68)	26.31 (0.69)	28.45 (0.73)	**29.31** (**0.79**)	29.22 (0.78)
	$\sigma = 40$	24.75 (0.59)	25.04 (0.60)	26.97 (0.65)	**27.98** (**0.73**)	27.72 (0.72)

Table 3. Denoising Results - PSNR (SSIM): Natural Images ($\sigma \in \{20, 30, 40\}$)

		LPG-PCA1	LPG-PCA2	BM3D1	BM3D2	NOLPP
Barbara	$\sigma = 20$	30.85 (0.86)	31.42 (0.89)	31.12 (0.88)	**31.74** (**0.90**)	30.46 (0.87)
	$\sigma = 30$	28.43 (0.77)	29.22 (0.85)	28.95 (0.82)	**29.81** (**0.86**)	28.32 (0.81)
	$\sigma = 40$	26.57 (0.68)	27.55 (0.81)	26.74 (0.75)	**27.98** (**0.82**)	26.81 (0.75)
Goldhill	$\sigma = 20$	30.06 (0.77)	30.22 (0.78)	30.40 (0.79)	**30.70** (**0.80**)	30.30 (0.79)
	$\sigma = 30$	28.02 (0.67)	28.51 (0.72)	28.67 (0.72)	**29.13** (**0.75**)	28.50 (0.72)
	$\sigma = 40$	26.38 (0.58)	27.23 (0.67)	27.21 (0.65)	**27.93** (**0.70**)	27.80 (0.66)
Lena	$\sigma = 20$	31.84 (0.81)	32.61 (0.87)	32.45 (0.85)	**32.98** (**0.87**)	31.91 (0.83)
	$\sigma = 30$	29.46 (0.71)	30.71 (0.83)	30.49 (0.80)	**31.25** (**0.84**)	29.91 (0.78)
	$\sigma = 40$	27.51 (0.61)	29.14 (0.80)	28.64 (0.74)	**29.79** (**0.81**)	28.43 (0.72)
Mandrill	$\sigma = 20$	26.52 (0.78)	26.42 (0.77)	26.28 (0.77)	26.60 (0.79)	**26.61** (**0.81**)
	$\sigma = 30$	24.26 (0.67)	24.20 (0.67)	24.22 (0.67)	24.56 (0.70)	**24.61** (**0.71**)
	$\sigma = 40$	22.93 (0.58)	22.92 (0.59)	22.44 (0.57)	23.11 (0.61)	**23.31** (**0.64**)
Stream	$\sigma = 20$	27.05 (0.77)	26.96 (0.77)	27.01 (0.77)	27.27 (0.78)	**27.30** (**0.79**)
	$\sigma = 30$	25.08 (0.67)	25.04 (0.67)	25.23 (0.68)	25.47 (0.69)	**25.50** (**0.71**)
	$\sigma = 40$	23.78 (0.58)	23.83 (0.59)	23.82 (0.59)	24.29 (0.62)	**24.34** (**0.64**)

range of similar patches is set to be $\sqrt{3\sigma^2 n^2}$ which is the widely used similarity measure for patches. Up to the specified distance, the patches are considered to be similar and are weighed accordingly, all other patches are weighed 0. Instead of learning different NOLPP basis for each overlapping patch, only one set of basis are learnt for windows of size 120×120 from the input noisy images of size 512×512. Each noisy patch is then projected on the corresponding NOLPP basis, and the coefficients are modified using the Weiner filter update.

The results of the proposed method are compared with some of the state of the art methods for denoising such as LPG-PCA [13] and BM3D [5]. Both the methods work in 2 stages, hence denoising performances in terms of PSNR and SSIM of both the stages are included in Table 2 and 3 along with the results of the proposed approach for some of the images from UCIC Texture database [2] and Lansel database of natural images respectively for 3 different noise levels i.e. $\sigma = 20,\ 30\ \&\ 40$. Proposed approach out performs both LPG-PCA and BM3D in most of the cases for texture images as these approaches tend to smooth out the fine textural details , whereas in case of natural images, proposed approach is comparable i.e. the finer details are still preserved but it under smoothes some homogeneous parts of the image (Figure 1).

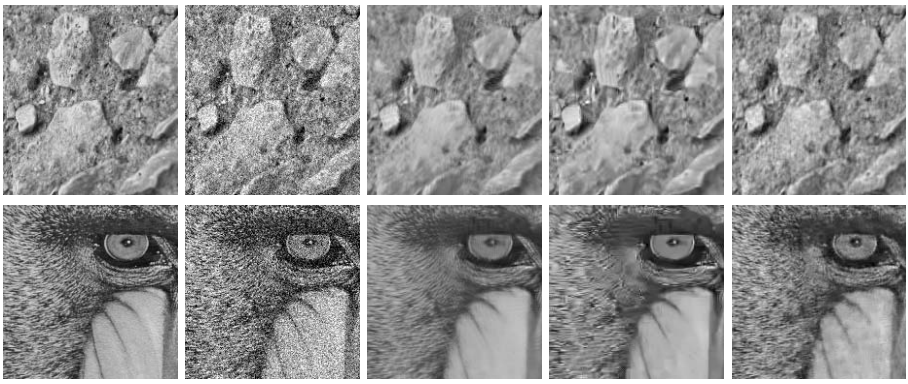


Fig. 1. From left to right, top to bottom: Original texture image, Noisy image with i.i.d. Gaussian noise ($\sigma = 30$), Results of denoising using LPG-PCA, BM3D, NOLPP. The same order for a natural image with i.i.d Gaussian noise ($\sigma = 40$) in the next row.

5 Conclusion

Locality Preserving Projection (LPP) though preserves the local structure, which could be essential for many applications, yet yields non-orthogonal basis that may not be acceptable for applications such as image denoising. This paper presented a method of obtaining orthogonal LPP basis. The application of such basis for pattern clustering and image denoising has been presented. It shows a strong reducibility capacity of data dimension which is effectively utilized in clustering. The denoising mechanism presented is almost outperforming most of the state of the art methods of denoising for textured images while comparable in case of the natural images. Denoising of natural images having some texture is appeared to be the best. This research will extend the applicability of New Orthogonal LPP towards other problems such as image deblurring and inpainting.

References

1. Mnist database of handwritten digits, http://yann.lecun.com/exdb/mnist/
2. Ucic texture database, http://wwwcvr.ai.uiuc.edu/ponce_grp/data/texture_database/samples/
3. Belhumeur, P., Hespanha, J., Kriengman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Pattern Analysis Machine Intelligence 19(13), 711–720 (1997)
4. Cai, D., He, X., Han, J., Zhang, H.: Orthogonal laplacianfaces for face recognition. IEEE Transactions on Image Processing 15(11), 3608–3614 (2006)
5. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3d transform domain collaborative filtering. IEEE Trans. on Image Processing 16(8), 2080–2095 (2007)
6. Dornaika, F., Assoum, A.: Enhanced and parameterless locality preserving projections for face recognition. Neurocomputing 99, 448–457 (2013)
7. Yu, J.W.G., Peng, H., Ma, Q.: Enhanced locality preserving projections using robust path based similarity. Neurocomputing 71, 598–605 (2011)
8. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23(6), 643–660 (2001)
9. He, X., Niyogi, P.: Locality preserving projections. In: NIPS (2003)
10. Hyvarinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley, New York (2001)
11. Jolliffe, I.: Principal Component Analysis. Springer, Berlin (2002)
12. Shikkenawis, G., Mitra, S.K.: Improving the locality preserving projection for dimensionality reduction. In: International Conference on Emerging Applications of Information Technology, pp. 161–164. IEEE Computer Society (2012)
13. Zhang, L., Dong, W., Zhang, D., Shi, G.: Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recognition 43(4), 1531–1549 (2010)

A Complex Diffusion Driven Approach for Removing Data-Dependent Multiplicative Noise

P. Jidesh[1] and A.A. Bini[2]

[1] Department of Mathematical and Computational Sciences
[2] Department of Electronics and Communication Engineering,
National Institute of Technology, Karnataka, India-575025
jidesh@nitk.ac.in, biniaa@gmail.com

Abstract. In this paper we propose a second-order non-linear PDE based on the complex diffusion function. The proposed method exhibits better restoration capability of ramp edges in comparison to other second-order methods discussed in the literature. The proposed model is designed for Gamma distributed multiplicative noise which commonly appears in Ultra Sound (US) and Synthetic Aperture Radar (SAR) images. The fidelity/reactive term augmented to the complex diffusive term is derived based on the Bayesian maximum a posteriori probability (MAP) estimator as detailed in Aubert and Ajol ([10]). The regularization parameter is selected based on the noise variance of the image and thus this adaptive method helps in restoring the images at various noise variances without manually fixing the parameter. The results shown in terms of both visual and qualitative measures demonstrate the capability of the model to restore images from their degraded observations.

Keywords: Complex diffusion, Multiplicative noise, Regularization method, Variational method.

1 Introduction

Restoration of images corrupted by data dependent multiplicative noise is a subject of practical interest in many imaging modalities. The data dependent multiplicative noise is commonly found in US and SAR images. Restoration under additive data dependent noise is quite extensively analyzed in the literature, the details can be found in Weickert ([3]) and the references therein. However, data dependent multiplicative noise is not so extensively analyzed in the literature. To our knowledge the first method of that kind was introduced by Rudin et al. ([5]). In this model the authors consider a data dependent multiplicative noise which follows a Gaussian distribution. This multiplicative Gaussian noise model is not of big practical interest. There are few dedicated studies those aim at restoring the images corrupted by Gamma and Poisson noise distributions, refer Aubert and Ajol ([10]) and Setzer et al.([12]) respectively, for further details. These restoration models discussed above use the Total Variation (TV) norm of the absolute gradient function as the functional to be minimized. The

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 284–289, 2013.

diffusion equation derived from the aforementioned functional gives a second order non-linear PDE. Quite a few second-order non-linear diffusion methods are also proposed for restoring images corrupted by the multiplicative noise, see Yu and Acton([4]), Sun et al.([7]) and Krissian et al.([9]). These PDE models are variations of the well known Parona-Malik ([1]) model, where the coefficient of diffusion function is modified to handle various noise distributions. However these models do not directly derive the evolution equations from the noise model.

2 Data-Dependent Multiplicative Noise Model

Multiplicative noise commonly appears in many imaging modalities, especially in medical and satellite imaging. A common multiplicative noise model assumed in various restoration problems is defines as:

$$u_0 = u \times n, \tag{1}$$

where n is a multiplicative noise which has a mean 1 and variance σ^2. The first kind of restoration method proposed for this noise model is Rudin et al. ([5]). In this work the authors assume a Gaussian multiplicative noise. The gradient descent solution to the PDE proposed in this model is:

$$u_t = div\frac{\nabla u}{|\nabla u|} - \lambda\frac{u_0^2}{u^3} - \mu\frac{u_0}{u^2}, \tag{2}$$

where div is a divergence operator. Here the two Lagrange multipliers λ and μ are dynamically updated to satisfy the constraints as explained in Rudin et al. ([5]). One of the major drawbacks of this model is that it assumes a Gaussian distribution for the noise. However, it is known from the history of image restoration that most of the imaging modalities of practical interest does not result in Gaussian distributed multiplicative noise. Here we assume a reflective boundary condition ($\partial u/\partial \boldsymbol{n} = 0$, where $\boldsymbol{n}$ is a unit vector normal to the plane) and the initial condition $u(x, y, 0) = u_0(x, y)$, where $u_0(x, y)$ is the initial image. We further note that these boundary and initial conditions are assumed for all the PDEs stated in this work, if not mentioned otherwise.

There are few works those consider the multiplicative noise model as a log translated additive model, refer Krissian et al.([8]) for details. Here the authors assume the data dependent nature of the noise in the log compressed images. However one can easily note that even in log compressed images the noise does not become completely independent of data. Therefore these models fail to provide an effective restoration under the multiplicative noise setup.

In another work Aubert and Ajol([10]) modified the reactive term of the model in order to incorporate a Gamma distributive nature of the noise. This modified model is quite applicable to various ultrasound and SAR images, where the noise amplitude is Gamma distributed or the image is speckled. The evolution PDE can be stated as:

$$u_t = div\left(\frac{\nabla u}{|\nabla u|}\right) - \lambda\frac{(u - u_0)}{u^2}. \tag{3}$$

Here λ is a regularization parameter and other symbols have their usual meanings.

There are some modifications suggested for the above cited work, in the recent literature, see Huang et al.([13]) and Huang et al.([11]) for details. In Huang et al. ([13]) the authors replace the TV regularizer with a Weberized TV regularizer for a better restoration. And in Huang et al.([11]) the authors provide a modified fitting term which is convex under all conditions, therefore the solution is unique.

Complex diffusion was introduced to image processing by Gilboa et al. in Gilboa et al. ([6]). A complex diffusion process motivated by the time dependent Schroedinger wave equation (a fundamental equation in quantum physics) can be written in an anisotropic manner as:

$$u_t = \nabla . c(Im(u)\nabla u), \tag{4}$$

where $Im(u)$ denotes the imaginary component of complex image u and

$$c(Im(u)) = \frac{e^{i\theta}}{1 + \left(\frac{Im(u)}{k\theta}\right)^2}, \tag{5}$$

where k is a threshold parameter. Further it was shown by the authors that the limiting case is $lim_{\theta->0}(Im/\theta) = t\Delta g_\sigma * u_0$, where $\Delta g_\sigma * u_0$ denotes the Gaussian convolved version of the Laplacian of the initial image. One can notice that since the coefficient of diffusion is a function of Laplacian of the image the ramp like edges are preserved more naturally in the filtered output, unlike other second-order non-linear diffusion methods.

3 The Proposed Model

In this work we propose a complex diffusion driven PDE for restoring specked images where the noise amplitude is Gamma distributed. In this proposed model the reactive term is borrowed from the multiplicative model proposed in Aubert and Ajol ([10]). This reactive term is derived with regard to the noise distribution using the Bayesian maximum a posteriori probability estimator. The proposed diffusion model looks like

$$u_t = \nabla . \left(c(Im(u))\nabla u\right) - \lambda \frac{(u - u_0)}{u^2}, \tag{6}$$

where the symbols have the same meanings as defined earlier and function $c(.)$ is defined in (5). Here λ is a regularizing parameter which controls the diffusive and reactive functionalities of the filter. We assume the limiting case of the diffusion coefficient at small values of θ. Since the negative diffusion leads to instability of the above PDE we assume the value of diffusion coefficient to be always positive. The initial condition $u(x, y, 0) = u_0(x, y)$ is assumed, where u_0 is the initial (observed) image which is real. One of the main features of complex diffusion is its capability to preserve ramp edges. It is observed that the gradient

based diffusion coefficient functions are not quite effective in detecting the main features of ramps; namely its end points. Moreover these functions have nearly uniform values across the whole smoothed ramp, causing a non-linear gradient-dependent diffusion to slow the diffusion process in those regions and therefore are less effective in noise reduction within the ramp-edges. However the Laplacian based diffusion coefficient functions have high magnitudes near the endpoints and low magnitudes elsewhere, and thus enables the nonlinear diffusion process to reduce noise over the ramp. Therefore, the proposed model is more effective in terms of noise reduction and ramp preservation compared to the other non-linear second-order diffusion processes like TV Rudin et al. ([2]) and Perona and Malik ([1]).

3.1 Numerical Implementation

We adopt an explicit time marching scheme for solving the proposed PDE numerically. The choice of the scheme is well justified by the reduced complexity and comparable accuracy. Here we show the discretization for the regularization term and the scheme is trivial for the reactive (fidelity) term.

$$u^{n+1} = u^n + \Delta t \left(\frac{\partial^-}{\partial x} \left(\frac{\partial^+ u}{\partial x} \times c(Im(u)) \right) + \frac{\partial^-}{\partial y} \left(\frac{\partial^+ u}{\partial y} \times c(Im(u)) \right) \right), \quad (7)$$

where $c(.)$ is defined in (5), u^n denotes the image at n^{th} iteration and the super script +/- denotes the forward/backward deference operators, respectively. The forward and backward difference operators along x and y are defined as:

$$u_x^+ = u(x+1, y) - u(x, y),\ u_x^- = u(x, y) - u(x-1, y),$$

$$u_y^+ = u(x, y+1) - u(x, y),\ u_y^- = u(x, y) - u(x, y-1).$$

Here we denote $\partial u / \partial x$ as u_x. Similarly we can derive $|\nabla u|$ as

$$|\nabla u| = \sqrt{\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2}. \quad (8)$$

With the help of above discretization we can discretize the PDE in (6)

4 Experimental Results and Discussions

We have tested the efficiency of the proposed method using various test images, however due to the page limitations we have demonstrated the result only for the image "woman". The input images are normalized to the range [0-1] in terms of their intensities. The Gamma noise is generated using a function written in Matlab. We have tested the images under various noise densities and the proposed model is found to be better in terms of noise reduction and details preservation as compared to the other methods considered in this work. A quantitative study is

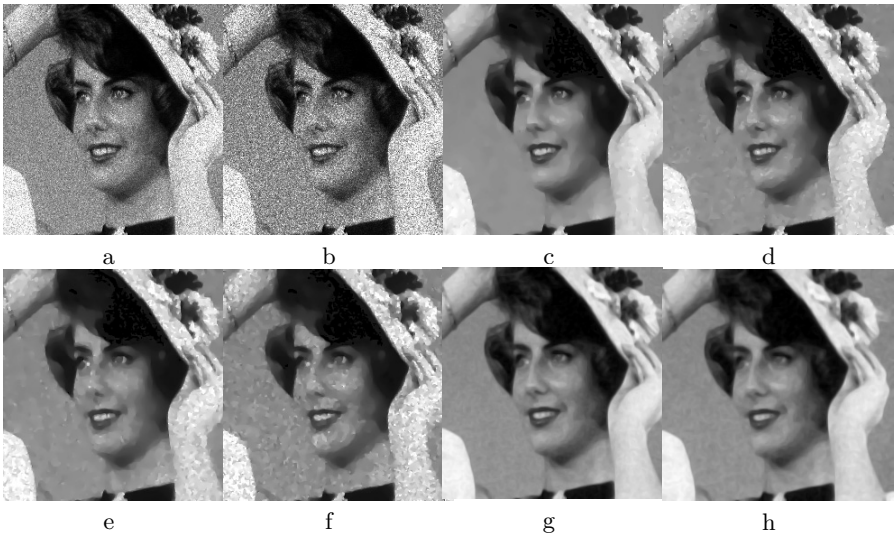


Fig. 1. Images restored using various methods: (a) and (b): Noisy images with mean 1 and noise variances 0.1 and 0.2, respectively; (c)and (d) restored using AA method; (e) and (f) restored using RLO filter; (g) and (h) restored using the proposed method

carried out using the statistical measure Peak Signal to noise Ratio (PSNR) and the results are provided in Table 1. Figure 1 shows the output of various methods considered in this work. The existing filters considered here for comparison are Aubert Ajol ([10]) filter and TV filter for multiplicative noise by Rudin et al. ([5]). From the results provided in Figure 1 one can confirm that the proposed method outperforms other methods in terms of visual appearance. Moreover the Peak Signal to Noise Ratio tabulated in Table 1, provides an understanding on the noise reduction capacity of various methods under consideration, tested with different images at a specified noise variance. The proposed method shows a better PSNR value compared to other considered methods. In addition to this the number of iterations taken by various methods to reach at the steady state was also analyzed in this study. We confirm that the proposed model takes slightly

Table 1. PSNR evaluated for various restoration methods for different images. (The gamma noise variance of the input image is :0.2).

Image Name	Noisy	RLO	AA	The proposed model
woman	11.27	12.91	15.27	16.63
cameraman	12.12	14.81	16.12	16.91
Lena	11.32	13.41	14.12	15.91
peppers	12.33	13.22	14.67	15.47

more number of iterations to reach the steady state in comparison to AA and RLO models, at different noise density levels.

5 Concluding Remarks

In this paper we have proposed a complex diffusion driven regularization filter for restoring images corrupted by multiplicative noise whose amplitude is Gamma distributed (speckled images). The restoration is found to preserve ramp features present in images. The experimental results prove beyond any doubts that the proposed model performs better in terms of visual and quantitative measures provided in the result section.

References

1. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 629–639 (1990)
2. Rudin, L.I., Osher, S., Fehtami, E.: Nonlinear total variation based noise removal algorithms. Physica D. 68, 259–268 (1992)
3. Weickert, J.: A Review of Nonlinear Diffusion Filtering Scale-Space, 3–28 (1997)
4. Yu, Y., Acton, S.T.: Speckle reducing anisotropic diffusion. IEEE Transactions on Image Processing 11, 1260–1270 (2002)
5. Rudin, L.I., Lions, P.L., Osher, S.: Multiplicative denoising and deblurring: theory and algorithms. In: Geometric Level Set Methods in Imaging, Vision, and Graphics, pp. 103–120. Springer, Berlin (2003)
6. Gilboa, G., Sochen, N.A., Zeevi, Y.V.: Image Enhancement and Denoising by Complex Diffusion Processes. IEEE Trans. Pattern Anal. Mach. Intell. 26, 1020–1036 (2004)
7. Sun, Q., Hossack, J.A., Tang, J., Acton, S.T.: Speckle reducing anisotropic diffusion for 3D ultrasound images. Computerized Medical Imaging and Graphics 28, 461–470 (2004)
8. Krissian, K., Kikinis, R., Westin, C.F., Vosburgh, K.: Speckle constrained filtering of ultrasound images. Proceedings of IEEE Comput. Vis. Pattern Recogn. (CVPR), 547–552 (2005)
9. Krissian, K., Westin, C.F., Kikinis, R., Vosburgh, K.G.: Oriented speckle reducing anisotropic diffusion. IEEE Transactions on Image Processing 16, 1412–1424 (2007)
10. Aubertt, G., Aujol, J.F.: A variational approach to removing multiplicative noise. SIAM Journal on Applied Mathematics 68, 925–946 (2008)
11. Huang, M., Ng, M.K., Wen, Y.: A New Total Variation Method for Multiplicative Noise Removal. SIAM J. Imaging Sciences 2, 22–40 (2009)
12. Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian Images by Split Bregman Techniques. J. Visual Communication and Image Representation 21, 193–199 (2010)
13. Huang, L., Xiao, L., Wei, Z.-H.: Multiplicative Noise Removal via a Novel Variational Model. EURASIP Journal on Image and Video Processing 2, 1–16 (2010)

Extracting Linear Features from SAR Images Using CGVF Snake Model and Beamlet Transform

V. Ramachandran and K. Vani

Department of Information Science and Technology
CEG, Anna University, Chennai, India
ramcndrn@gmail.com,
vani@annauniv.edu

Abstract. Extracting Linear Features from Microwave Images (SAR) using CGVF Snake Model and Beamlet Transform is proposed in this paper. Microwave images are independent of climate. In certain durations these images are affected by noise. So noise is removed in this work. For CGVF Snake model, edges are detected and GVF field is produced followed by Snake is initialized and then two external constraint forces are developed. The first one will points on the snake and determine the basic shape of a snake. The second one generating the curves is smooth and grows in a correct direction. For each iteration, the snake is deformed in edges and removes the discontinuities for extracting the linear features. For beamlet Transform, initially the SAR image is enhanced by applying the Non-uniform Background elimination method followed by Beamlet Transform based algorithm was applied. This algorithm recursively partitions the image into sub-squares to build a beamlet dictionary to perform the transform followed by discontinuities are linked to extract the linear features. The results of feature extraction from microwave images using CGVF Snake model and Beamlet Transform to increase the correctness and quality of satellite Mapping.

Keywords: gradient vector flow (GVF), synthetic aperture radar (SAR), controllable gradient vector flow (CGVF), Beamlet Transform, linear feature extraction.

1 Introduction

Microwave images(SAR) are capable of operating under weather conditions and have more noise(i.e. Salt and pepper noise and also speckle noise) .These images are suitable options for most of the real time application such as road extraction, ridge extraction etc. Due to noisy nature, SAR image is not a suitable option for real time application, so proper pre-processing is mandatory option for SAR image.

Removing noise in the image needs the filter technique such as Median filter [7]. Median filter helps to reduce the conflict occurs due to presence of salt and pepper noise and the speckle noise is to be reduced by thresholding technique. Because of

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 290–303, 2013.

curve Nature, Most of the edge detection technique will not suitable for extracting the edges of the road.

After removing the noise, edges are detected. Normally the Edge in an image may be defined as a rapid photometric change in the registered intensity values. From a pixel level perspective, edge can be viewed as the regions of an image where the image values undergo a sharp variation. In discrete 2-D image edge is detected as linear combination of the point's singularities between pixels. The edge detection is discussed in [3].In proposed work canny edge detection is applied to detect the edges followed by feature is extracted.

Microwave images contain many linear features (i.e. road and ridge networks). Extracting those linear features is a fundamental task in many image processing and computer vision applications. Traditionally the snakes are developed to detect features. In [1], M. Kass and A. Witkin presents Snake is a curve defined within an image domain that can move under the influence of internal forces coming from the curves themselves and external forces computed from the image data. In [2], Xu et al. present a class of external force named gradient vector flow (GVF) that largely solves the problems such as initialization and poor convergence to boundary concavities of traditional snakes. In particular, they may be not pulled to the target feature correctly.

Aiming at improving the performance of traditional GVF snakes, ***the improved CGVF snakes are employed in this work to close gaps in linear feature extraction*** which is important step involved. ***Further improve the feature extraction process beamlet transform is proposed*** in this paper. In [11], XiaoYu DOU HongXun SONG presents the beamlet Transform based algorithm to detect the cracks in image. also E.Salari and Y.Zhu [7] presents the road extraction using the Beamlet transform. But they also contain some discontinuities. But they also contain some discontinuities.

In proposed Beamlet Transform process, the image is be enhanced [6] followed by recursively partitioned into sub-squares to build the beamlet dictionary and find the beamlet Transform co-efficient to construct the beamlet image. The discontinuities present in the beamlet image are linked efficiently by using Thinning method. Finally compare the efficiency of CGVF Snake Model and Beamlet Transform method to find the best approach for extracting the linear feature in efficient manner.

The rest of the paper is organized as follows Section 2 contains the system description. It describes the overall system contains noise removal, edge detection and linear Feature Extraction using CGVF Snake Model, Image Enhancement, Linear Feature Extraction Using Beamlet Transform and Performance evaluation. Section 3 discusses experimental results and finally Conclusion is given in Section 6.

2 System Description

The System Architecture for linear feature extraction from microwave images using CGVF Snake Model and Beamlet Transform is illustrated in figure1.

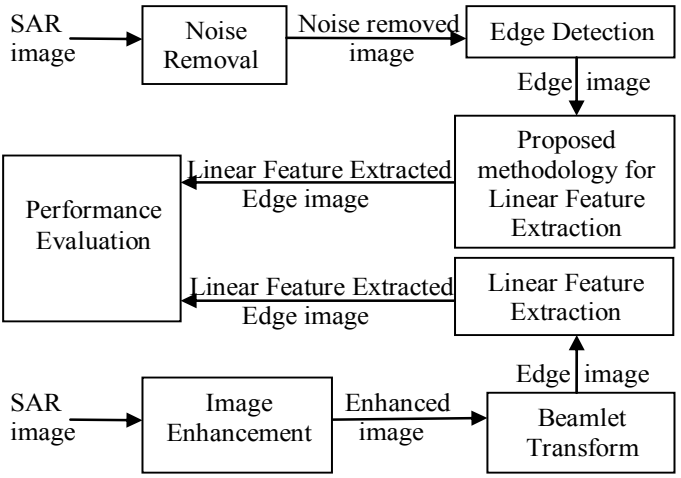


Fig. 1. System Architecture

Each process of the system is to be constructed and described below.

A. Noise removal

Presence of noise in an image is the major drawback for SAR image. It shows false edges in the Edge mapped image. So keeping Noise Removal process prior to all is an effective idea is illustrated in figure 2.

Initially Noisy (SAR) image is converted to gray scale image is followed by Histogram equalization to enhance the contrast and median filter helps to remove the salt and pepper noise and finally produces the noise removed image.

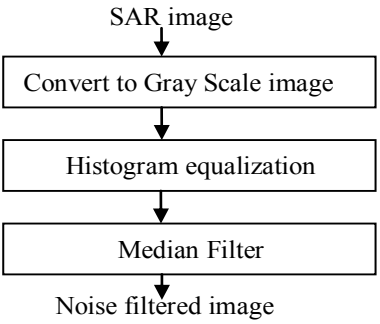


Fig. 2. Noise Removal

1) *Convert to Gray Scale image*: It converts RGB to Gray scale by replacing the RGB values.

$$Y = wR\boldsymbol{R} + wG\boldsymbol{G} + wB\boldsymbol{B} \quad (1)$$

Where **R, G, B** and Y are Red, Green, Blue and Gray pixel . wR, wG, wB are Weight Parameters their values are wR =0.2125, wG =0.7154, wB= 0.072

2) Histogram equalization: It is equalize the Histogram to enhance the contrast of the image. It has following steps.

a) Form the cumulative Histogram.(Cum)

b) Normalize the value by dividing it by Total number of pixels. (Cum/Total)

c) Multiply these values by the maximum gray level(L) value and round off the value.(cum/total)*(L-1)

d) Map the original value to the result of step a one-to-one correspondence.

3) *Median Filter:* It is used to remove the Salt and pepper Noise. It has following steps.

a) Read the pixel values.

b) Sort the pixel values.

c) Find Median value.

d) Substitute median value to central values for each mask.

The above mentioned processes are help to remove the noise from SAR images efficiently and produce the noise filtered SAR image as a output.

B. Edge Detection

Canny Edge detection [4] is used in this process. Initially noise filtered images was smoothed followed by compute the edge map gradient. To find the Sharp edges, Non-maximum suppression is used. Double thresholding helps to extract the strong and weak edges. To remove those weak edges hysteresis is applied and finally produces the edge image. These processes are illustrated in the following figure 3.

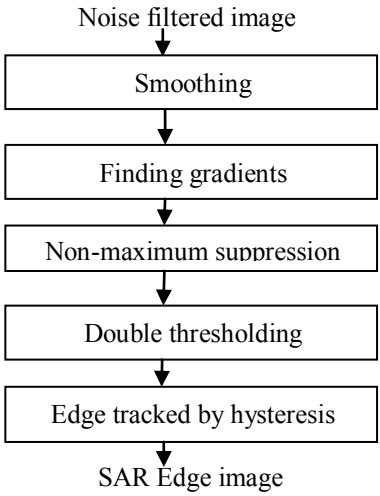


Fig. 3. Edge Detection

1) *Smoothing*: The image is smoothed by following equation.

$$g(m,n)= G_\sigma(m,n)*f(m,n) \quad (2)$$

Here $G_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{m^2+n^2}{2\sigma^2}]$,where σ=Standard deviation. Here σ=1.4,f(m,n) is noise filtered image.

2) *Finding gradients:* Gradient magnitude can be determined by the following formula.

$$M(n,n) = \sqrt{g_m^2(m,n) + g_n^2(m,n)} \quad (3)$$

Here g_m, g_n are gradient in x and y direction.To finds the direction of the edges by using the following Formula.

$$\theta(m,n) = tan^{-1}[g_n(m,n)/g_m(m,n)] \quad (4)$$

3) *Non-maximum suppression:* It is used to convert the gradient magnitudes to "sharp" edges. It has the following process.

a) Compare the edge strength of the current pixel with the edge strength of the pixel in the positive and negative gradient direction

b) If the edge strength of the current pixel is largest; preserve the value of the edge strength. If not remove the value

4) *Double thresholding:* It is used for finding Strong and Weak edges. It has the following steps.

a) Edge pixels stronger than the high threshold are marked as strong. M(x,y)=M(x,y)>T1→Strong edge.

Here T1 is high threshold value. Where T1=0.8.

b) Edge pixels weaker than the low threshold are suppressed. M(x,y)=M(x,y)>T2→removed edge.

Here T2 is low threshold value.WhereT2=0.2.

c) Edge pixels between the two thresholds are marked as weak. M(x,y)=T1<M(x,y)>T2→Weak edge.

5) *Edge tracked by hysteresis*: It is used to eliminate weak edges and ensure the strong edges. It has the following steps.

a) The edge pixels are divided into 8-connected Neighbourhood.

b) It's containing at least one strong edge pixel are then preserved, while other are suppressed.

The above processes are helps to produce an edge image as an output. These edge images also contain some unwanted features. So to extract those linear features (i.e. roads, ridges) CGVF Snake model is proposed in this paper.

C. Proposed methodology for linear feature extraction

To extract the linear feature is difficult process in many existing algorithm such as Snakes [1] and GVF Snake [2] algorithm. In particular, they may be not pulled to the

target feature correctly. So the CGVF Snake model is proposed to improve existing algorithms for extracting linear features and remove the discontinuities efficiently.

Initially GVF field is produced (i.e. internal and external forces of an image) followed by the Snake is created and then CGVF Snake [5] is employed to close the gap (i.e. remove the discontinuities) and finally produces the linear features extracted image. These processes are illustrated in Figure 4.

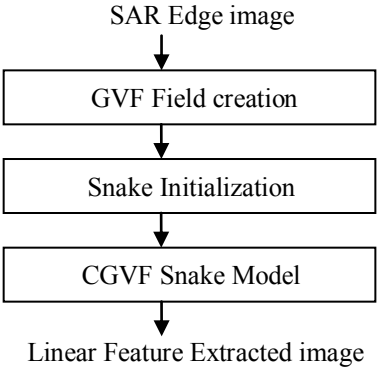


Fig. 4. Linear Feature extraction

1) *GVF Field creation*: It is defined as the vector field V=(u,v). This vector field is created by the following equation.

$$\mu\Delta^2 u-(u-f_x)(f_x^2+f_y^2)=0$$
$$\mu\Delta^2 v-(v-f_x)(f_x^2+f_y^2)=0 \quad (5)$$

Where f_x and f_y are partial derivative of edge image. Δ^2 is Laplacian operator. $\boldsymbol{\mu}$ is used to find the smooth of force field .

2) Snake Initialization: snakes are computer-generated curves that move within images to find object boundaries.

$$c(s) = (x(s), y(s)), s[0,1] . \quad (6)$$

Its energy functional can be written as

$$E_{snake} = \int_0^1 [E_{\text{int}}(\boldsymbol{c}(s)) + E_{image}(\boldsymbol{c}(s)) + E_{con}(\boldsymbol{c}(s))]ds \quad (7)$$

Here E_{int} - internal energy of the snake, E_{image} - image forces, E_{con} -external constraint forces.

After initial Snake created then it is moved in image through internal energy, external energy and image force.

3) *CGVF Snake Model (proposed method):* It is an improved GVF snake model. It is developed from *GVF* Snake [2]**. *In GVF snake, snakes are moved in undesired path in edges. It's also generating ears. So the CGVF Snakes are employed to remove the ears and passed through edges in correct direction.*** It is developed by the following external constrains forces created in this paper.

a) *Point pin energy:* It is fixes specified points P on the Snake. Collection of these connected points to form a basic shape of the Snake. it's energy function can be written as

$$E_{pin} = \kappa \sum_{p \in P} \prod_{c_i \in c} H(\|c_i - p\| - \delta) \tag{8}$$

Here κ-weight parameter, δ-is the tolerance determining whether two points are the same or not and c_i - Snake points and H -Heaviside function is determined by

$$H(z) = \begin{cases} 1, \; if \;\; z \geq 0 \\ 0, \; else \end{cases} \tag{9}$$

b) *Ear avoiding energy:* it makes the curve smoother and grows in the correct direction. it's energy function can be written as

$$E_{ear} = \rho[1 - \prod_{c_i, c_j \in c, i \neq j} H(\|c_i - c_j\| - \delta)] \tag{10}$$

Here ρ- Weighted Parameter.c_i and c_j are start and end points of snake.

External forces are formed by $E_{con} = E_{pin} + E_{ear}$ (11)

This E_{con} pixels values are subtituted in equation (7) for snake deformation.

These above mentioned processes are used for extract the linear features. ***For each itration the snake is grow in edges correct direction and close the gap(i.e. remove the discontinuities) to extract the linear features efficiently***. Further improving the feature extraction from SAR image Beamlet Transform also presented in this paper.

D. Image Enhancement

To improve the background lighting variation of the SAR image, Image Enhancement [6] process was applied. It helps to eliminate the Non-Uniform background of image and separate features from background thresholding were applied. These processes are illustrated in Figure 5.

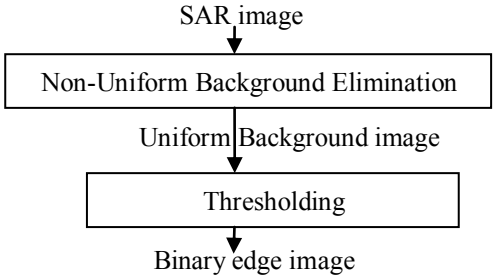


Fig. 5. Image Enhancement

1) *Non-Uniform Background Elimination:* the uniform background of image was produced by calculating the multiplication factors that eliminate the background lighting variations. It has the following steps

a) Divide the image into rectangular windows. The image base is 256*256, and the window size is set to 16*16 pixels.

b) Calculate the mean, minimum, and maximum gray level of each window. Denote them as G_{mean} , G_{min} and G_{max} .

c) Set an upper limit and a lower limit, for which the points with gray levels outside the limitations are considered as suspicious points for noise, crack, or other objects on the road.

$$r_l=G_{mean}-(G_{mean}-G_{min})*30\%$$
$$r_h= G_{mean}+(G_{max}-G_{mean})*30\% \quad (12)$$

d) With the exemption of the suspicious points, recalculate the mean value of the gray level G_{mean} ' .

e) The amplitude factor correction is calculated as $f = B /G_{mean}$, thus the modified picture intensity value is defined as I'=I*f.

2) *Thresholding:* It is used to separate features(i.e. road or crack edge pixels)from uniform background image. It is defined as

$$T=Mean-(Mean-min)*50\%. \quad (13)$$

If gray value I(x,y)<T then set value I(x,y)=0 else set valueI(x,y)=1;

These above process are helps to enhance the image and produce the binary edge image.

E. Beamlet Transform (proposed method)

The binary edge image contains linear or curvilinear features. These features are extracted by Beamlet Transform based algorithms were proposed in this paper. The JBEAM tool [8] was performs the Beamlet Transform operation.

Initially the enhanced image was partitioned followed by beamlet dictionary (i.e. tree based Beamlet Representation) is formed by applying the beamlet coder [8] and each beamlets are normalized. thresholding and testing was remove the undefined beamlets and produces the beamlet edge image. These processes are illustrated in Figure 6.

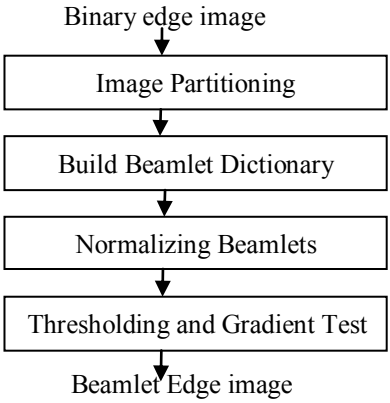


Fig. 6. Beamlet Transform

1) *Image Partitioning:* The first stage is partitioning the image into dyadic squares, marking the boundaries of squares with equi-spaced vertices at pixel spacing.

2) *Build Beamlet Dictionary:* The beamlet dictionary is built by connecting pairs of vertices within a dyadic square. Beamlet coder was build the tree based beamlet representation (BR).here the length and orientation of beamlet was calculated by converting BR into Symbol Stream and reconstructs the symbol stream into Beamlets.

3) *Normalizing Beamlets* :the beamlets are normalized by the following equation

$$T_f^{'}(b) = T_f(b) / \sqrt{L(b)} \tag{14}$$

Here $T_f(b)$-beamlet co-efficient ,L(b)-length of the beamlet.

4) *Thresholding and Gradient Test:* The normalized beamlets are compared with threshold value is given by

$$T = [\max(T_f^{'}(b)) - \min(T_f^{'}(b))]/2 \tag{15}$$

If $T'_f(b)$ is greater than T, a line segment is considered to exist in the sub-square, if $T'_f(b)$ is less than the threshold value, the beamlet is discarded. Finally test whether all sub-square are connected or not.

These above process are helps to produces the beamlet edge image. These images also contain some discontinuities.

F. Linear Feature Extraction

After the Beamlet Transform, beamlet edge image contains some discontinuities .***To extract the linear features and remove discontinuities post processing algorithm was developed in this paper.*** The post processing includes edge linking and thinning process [9] was illustrated in Figure 7.

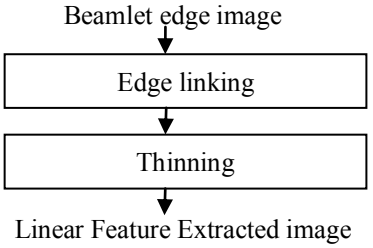


Fig. 7. Linear Feature Extraction

1) Edge Linking: It is used to link the discontinuity edges. It has the following steps.

a) Find the gradient in x and y direction (i.e G_x and G_y).

b) Edge linking is performed by the following equation

$$I(x,y)=sqrt(G_x^2+G_y^2) \tag{16}$$

Here I(x,y)-edge linked beamlet image.

2) Thinning: It is used to remove the unwanted beamlets and produces the linear feature extracted image.

These above process are helps to extract the linear features efficiently and performances of the images are evaluated in this paper.

G. Performance Evaluation

The performance [10] of the image is calculated by the following process.

1) *Peak Signal to Noise Ratio(PSNR):*The PSNR is used to find the difference between two images. It is defined as

$$PSNR = 20 * log10(b/MSD) \tag{17}$$

Where b is the largest value of the signal (typically 255or 1) and MSD is mean square difference between two images.

$$MSD = \frac{\sum_{r,c} (I_s(r,c) - I_d(r,c))^2}{R \times C} \tag{18}$$

Where $I_s(r,c)$- noisy image. $I_d(r,c)$-denoised image. R, C - number of rows and columns respectively.

2) *Noise Mean Value (NMV)*: It determines the quantity of noise content in the denoised image. Lower value means lower noise content. It is defined as

$$NMV = \frac{\sum_{r,c} I_d(r,c)}{R \times C} \tag{19}$$

3) Noise variance (NV): It determines the content of noise in the image. The lower variance gives cleaner image as more noise is reduced.

$$NV = \frac{\sum_{r,c} (I_d(r,c) - NMV)^2}{R \times C} \tag{20}$$

3 Experimental Results

The proposed CGVF Snake model has been implemented and the performance results are presented. Figure(8) shows an TRA_SAR_image1 represents the road netwok and Figure (9) shows an TRA_SAR_image2 represents the montain area.These images are taken from china with high resolution and high frequency.

Fig. 8.

Fig. 9.

During the Noise Removal process , Noise removed image is produced. Figure(10) and Figure(11) shows an noise removed image1&2 for TRA_SAR_IMAGE1 and TRA_SAR_IMAGE2.Here size of images are 256×267 pixel and [3×3] median filter is used in this process.

Fig. 10.

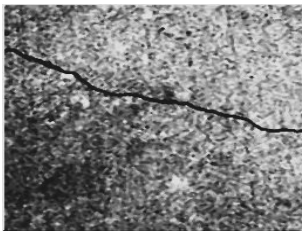


Fig. 11.

During the edge detection process,edge image is produced. Figure(12) and (13) shows an Edge image1&2.

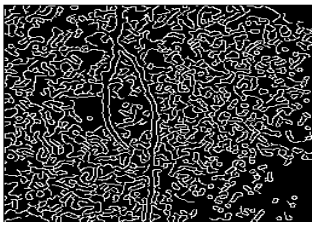


Fig. 12.

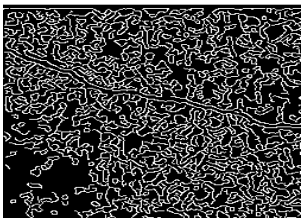


Fig. 13.

During linear feature extractoion process,linear feature extracted image is produced.Figure(14) and (15) shows an Linear Feature Extracted image1&2.for GVF creation process , μ is set to be 0.2 and iteration=80.

Fig. 14.

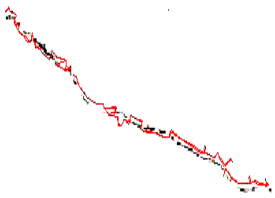


Fig. 15.

Linear Feature Extraction based on Beamlet Transform has been implemented and performance results are presented. During Edge Enhancement process, binary edge image is produced. Figure (16) and Figure (17) shows binary edge image1&2.here threshold T=0.43 for binary edge image1 and T=0.45 for binary edge image2.

Fig. 16.

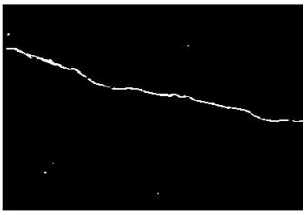


Fig. 17.

During Beamlet Transform process, Beamlet edge image is produced. Figure (18) and Figure (19) shows Beamlet edge image1&2.size of images should be 256×256 pixels.

Fig. 18.

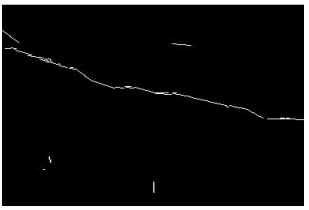


Fig. 19.

During Linear Feature Extraction process, linear feature extracted image is produced. Figure (20) and Figure (21) shows linear Feature extracted image1&2. Size of images should be 256×256 pixels.

Fig. 20.

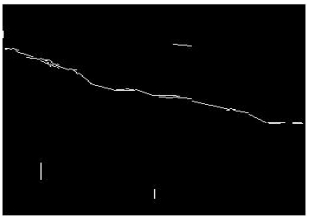


Fig. 21.

The Performance of the CGVF Snake Model and beamlet transform were evaluated and corresponding results are Shown in Table I and Table II.

Table 1. Performance for CGVF Snake Model

Figure Name	NMV	NV	MSD	M	Itera Tion	PSNR (db)
SAR_ IMAGE1	125.55	4851.68	5729.7	0.2	100	10.583
SAR_ IMAGE2	127.06	4669.35	3360.3	0.2	100	12.90

Table 2. Performance for Beamlet Transform

Figure Name	NMV	NV	MSD	T	No of Nonzeros	PSNR (db)
SAR_ IMAGE1	83.910	76.456	123.877	0.42	944	27.234
SAR_ IMAGE2	112.599	112.599	363.552	0.45	707	22.559

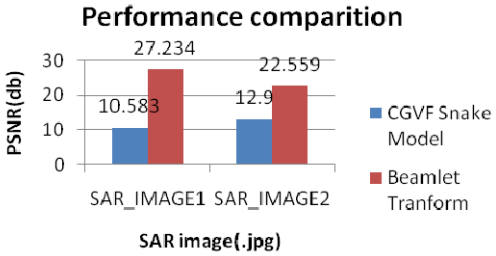


Fig. 22. BAR Graph(performance comparition)

Figure. 22 shows the BAR Graph.Here the PSNR value for Beamlet Transform was high compare to the CGVF Snake Model.Hence, from the BAR Graph, the performance of the beamlet Transform based algorithm gives best results compared to the CGVF Snake model.

4 Conclusions

Linear Feature is an important characteristic in Microwave images. These images are affected by noise. So the noise is removed in this work by using median filter method. After that the edges are detected efficiently by applied a canny edge detection algorithm.GVF field is produced and then Snake was deformed in edges by applying an external constrain forces of CGVF Snake model. After Snake deformation was completed, the linear features are extracted. Non-uniform background elimination process helps to enhance the image followed by applied the beamlet transform to extract linear features and discontinuities are removed in this paper. Finally CGVF Snake Model and Beamlet Transform are compared and experimental results are presented. From the experimental results, beamlet transform is the best approach for extracting the linear feature efficiently and increase quality of satellite Mapping. In future work, the new algorithm has to be implemented for improving the extraction of linear feature from microwave images to overcome the problems of missing edges in this work.

References

[1] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. International Journal of Computer Vision 1, 321–331 (1987)
[2] Xu, C., Price, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7, 359–369 (1998)
[3] Min, W., Yanning, Z.: Gray-scale Image Edge Detection Based on directional Beamlet Transform. IEEE Transactions on Image Processing 7, 1120–1124 (2008)
[4] Canny Edge Detection (May 23, 2009), http://www.cse.iitd.ernet.in/~pkalra/csl783/canny.pdf
[5] Deng, X.-P., He, C., Sun, H.: An Improved GVF Snake Model and its Application to Linear Feature Extraction from SAR Images. IEEE. ICSP, 2063–2067 (2010)
[6] Ying, L., Salari, E.: Beamlet Transform Based Technique for Pavement Image Processing and Classification. IEEE Transactions on Image Processing, 141–145 (2009)
[7] Salari, E., Zhu, Y.: A Road Extraction Method Using Beamlet Transform. IEEE Transactions on Image Processing, 1130–1134 (2012)
[8] Huo, X., Chen, J.: JBEAM: multi scale curve coding via beamlets. IEEE Trans. Image Processing 14, 1665–1677 (2005)
[9] Edge Thinning Used in the SUSAN Edge Detector, http://users.fmrib.ox.ac.uk/~steve/susan/thinning/thinning.html
[10] Joshi, R., Garg, R.D.: Pre-processing of TerraSAR-X Data for Speckle Removal: An Approach for Performance Evaluation. Indian Soc. Remote Sens. 40(3), 371–377 (2012)
[11] Wei, N., Zhao, X., Dou, X., Song, H.: Beamlet Transform Based Pavement Image Crack detection. In: IEEE International Conference on Intelligent Computation Technology and Automation, vol. 13, pp. 881–883 (2010)

Learning Semantic Interaction among Graspable Objects

Swagatika Panda, A.H. Abdul Hafez, and C.V. Jawahar

International Institute of Information Technology, Hyderabad, India

Abstract. In this work, we aim at understanding semantic interaction among graspable objects in both direct and indirect physical contact for robotic manipulation tasks. Given an object of interest, its support relationship with other graspable objects is inferred hierarchically. The support relationship is used to predict the "support order" or the order in which the surrounding objects need to be removed in order to manipulate the target object. We believe, this can extend the scope of robotic manipulation tasks to typical clutter involving physical contact, overlap and objects of generic shapes and sizes. We have created an RGBD dataset consisting of various objects present in clutter using Kinect. We conducted our experimentation and analysed the performance of our work on the images from the same dataset.

Keywords: Robotic Vision, Support Relation, Support Order, Semantic Interaction, RGBD.

1 Introduction

Understanding semantic interaction among the objects plays an important role in various robotic manipulation tasks in clutter. However, most of the robotic grasping and manipulation tasks remain confined to isolated objects, and often lying on planar surfaces [2, 8]. Recently, there has been a few attempts to infer pairwise support relationship among objects in clutter. Rosman *et al.* [9] predict three types of support relations: "on", "adjacent to", "both adjacent and on" using Kernel SVM. However, they assume that background is already segmented from the data. Sjöö and Jensfelt [11] infer four types of relations among each pair of object, viz. casual support, support force, protection and constraint using logistic regression classifier. But they work on simulated environment and hence remain limited to the inherent imperfections of simulation. Silberman *et al.* [10] perform inter-class support inference among regions of four major structure classes, viz. floor, wall, furniture and props using linear programming. They work in cluttered indoor settings. However, their work does not exploit the spatial relationship among different objects that overlap onto each other. Unlike the previous works, the aim of our work is to perform support inference in clutter involving both graspable and non-graspable objects, infer both direct and indirect support relations and then predict the "support

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 304–312, 2013.

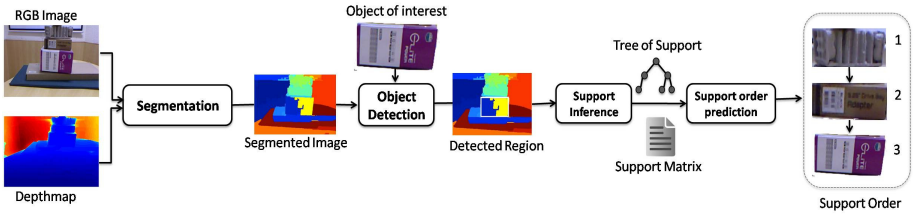


Fig. 1. Overview of our framework: Given the RGB and depth images and target object, our algorithm predicts the Support Matrix and the Support Order of the objects

order" in which the supported objects should be removed so that the object of interest can be manipulated without causing damage to the environment.

In our proposed framework, we learn the semantic interaction among the objects in clutter in three phases. At first, the regions corresponding to the graspable objects are separated out from the background entities in a cluttered indoor scene. Then, a target object is detected in clutter and its support relationship w.r.t. other objects is inferred hierarchically. Four kinds of support relations, i.e., "support from below", "support from side", "containment" or "none" are inferred. A "tree of support" is simultaneously built during support inference to encode the support relationship. Finally, the tree of support is traversed to predict the "support order". During tree traversal, special scenarios are identified and addressed so that minimal damage occurs when the objects are removed according to the predicted support order. We use different geometric features using both RGB and depth information to effectively capture the physical properties of objects. We validate the inferred support relationship and support order on various images from our RGBD dataset captured using Kinect in different indoor settings. In contrast to our previous work [7], where support relationship among every pair of objects present in the scene is obtained prior to detecting the target object and performing support order prediction, in the proposed work, support inference for all pairs of objects is avoided by posing it as an object-centric task. Here, support inference is performed hierarchically w.r.t. the object of interest. Using this approach, the number of comparisons required is reduced from $O(n^2)$ to $O(nlog(n))$. We believe, the predicted support order can be very useful in many applications such as grasping and object manipulation tasks in clutter involving overlaps and support by multiple objects. The overall framework of our work is explained through the block diagram shown in Fig. 1.

2 Support Inference

2.1 Segmentation and Object Detection

The image is segmented into different regions and target object is detected. Hierarchical segmentation method of Hoiem *et al.* [4] is used to segment the scene. This method uses superpixels segmented using Arbelaez's method [1]. Both 2D and 3D features of images are used for segmentation. The segmented regions are

provided as input to the object detection and support inference modules. In the object detection module, SIFT feature matching [3, 6, 12] between the template image of the object of interest and the input image is performed. The outliers are discarded by applying RANSAC. The segmented regions corresponding to the matched points of the input image are merged into one region and chosen as the region corresponding to object of interest O, i.e., the object to be grasped. This approach ensures that the entire object region is chosen for grasping. Given the region corresponding to the target object, support relationship among different regions is predicted hierarchically.

2.2 Learning and Inference

A cascade of classifiers is applied to the segmented regions and the target object for support inference. The learning methods, classifiers and approach for inference is discussed in detail below.

Structure Class Classification. An indoor environment typically consists of entities with distinct structural properties such as floor, walls as well as entities with highly diverse structural properties such as furniture. Given a cluttered scene, it is important to first determine which objects are graspable and segregate them from other non-graspable entities. A logistic regression classifier is trained to predict the structure class (floor/wall/furniture/graspable objects) [10] of each region. The regions predicted as graspable objects are selected for support inference while other regions are discarded as background. By discarding regions corresponding to walls, floor and furniture, unnecessary comparisons between all the regions in the image are avoided. In our experimentation, logistic regression is performed using stochastic gradient descent algorithm over the normalized features. The values of learning rate, batch size and maximum number of updates are empirically set to 10^{-5}, 100 and 7000, respectively.

Support Class Classification. In clutter where objects overlap on one another, an object supports multiple other objects. In order to access any object O without causing damage to the environment, all objects that O supports must be identified and removed. Our goal is to predict these supported objects. Pairwise support relationship among these objects is inferred using a 3-layer feed-forward neural network based support classifier. We perform support inference hierarchically instead of comparing all regions with each other for efficiency. Given a pair of regions (A, B), the support class classifier predicts if B supports A "from below", "from side", "contains" it or "not related" to it.

The support features for the regions corresponding to only the graspable objects are extracted and normalized for training the neural network. Sigmoidal activation function is used for the four output units to limit all the outputs to a fixed range([0 1]) so that the outputs can be interpreted as probabilities. The number of hidden nodes is kept as one tenth of the size of the training data to avoid over-fitting. Stochastic gradient descent algorithm is used to minimize the

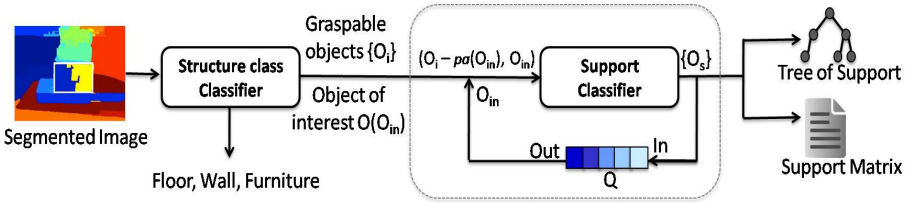


Fig. 2. Support Inference Module: The graspable object regions are filtered using structure class classifier. Support inference is performed hierarchically. tree of support and Support Matrix are generated as final output.

cross-entropy loss function which is appropriate for dealing with probabilities. The most probable support class is assigned to the pair of regions given as input. In our experimentation, the values of number of hidden nodes, learning rate, batch size and maximum number of updates are empirically set to 50, 10^{-4}, 100 and 10,000, respectively.

Hierarchical Support Inference. An object can support multiple objects both directly and indirectly through other objects it directly supports. To ensure minimal damage during manipulation, support inference is performed hierarchically beginning with the target object.

Suppose, $S = \{O_i | i = 1, \ldots, n\}$ is the set of n graspable objects/regions in the scene and $O \in S$ is the target object. A tree of support T is built with the target object in the root to encode the predicted support relationship. (The details of traversal of T are discussed in Section 3). Instead of inferring for all possible pairs of objects which will require $n(n-1)/2$ comparisons, support relations of all the objects that O directly supports are inferred at first. Every object is paired with O and support relations of each of the pairs $\{(O_i, O) | O_i \in S - \{O\}\}$ are inferred from the support classifier. The objects $S_s = \{O_s\}$ predicted as objects supported by O are stored in a FIFO queue Q. They are also inserted into the tree T as child nodes of O. The object pulled from the queue Q is again fed to the support classifier to predict the objects that it supports in turn. All other regions except the supporting regions(s) or parent(s) $pa(O_s)$ and grandparents $pa(pa(O_s))$ of O_s, are paired with O_s for the prediction. Note that special care is taken to discard all the parents and grandparents of the object O_s in order to avoid loops which may lead to damage in practical scenario. The support inference is performed hierarchically until all the outcomes of the classifier are negative, i.e., all the support relationships are predicted as "not related" and consequently the queue Q is empty.

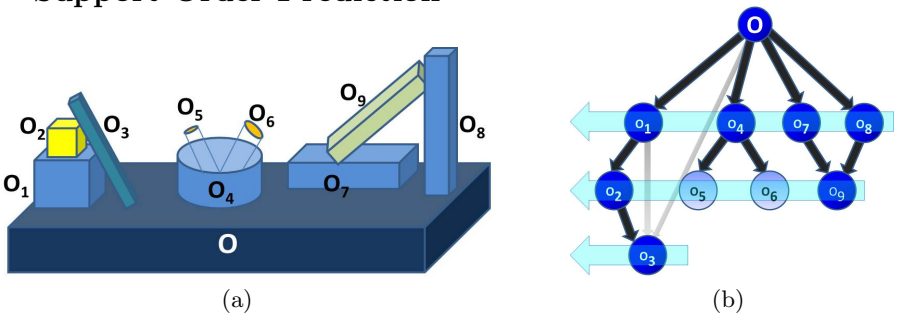
The advantage of performing support inference hierarchically instead of inferring for all pairs of objects is that this approach significantly reduces the number of comparisons required from $O(n^2)$ to $O(nlogn)$. Fig. 2 illustrates the process of support inference.

2.3 Features

For structure class classification, features proposed by Silberman *et al.* [10] are used. The features include SIFT features, histograms of surface normals, 2D and 3D bounding box dimensions, color histograms and relative depth.

For support classifiers, various features that capture the pairwise relationship of regions are used. Support feature $f(A, B)$ determine if region A is supported by B, but not the reverse. Hence, these features are not symmetric in nature. We adapt support features from [10]. The features include SIFT features of both regions to capture their individual characteristics, location features that capture absolute 3D positions of the regions and geometric features that capture different geometric relations between the two regions. In addition to this, we introduce five more new geometric features for improvement in encoding the pairwise relationships briefly discussed below. (i) **Proximity** between two objects $f_p(A, B)$ is measured as the ratio of distance in 3D between the centroids of the two objects and the sum of radii of the spheres circumscribing the two regions. The radius of the circumscribing sphere is approximated as the distance between the centroid and the point farthest from the centroid. (ii) The amount of overlap $f_{br}(A, B)$ is measured by **Boundary Ratio**, i.e., the ratio of the length of the boundary of the supported object A in contact with B with the perimeter of A.(iii) **Depth Boundary** $f_d(A, B)$ is used to differentiate visual occlusion and actual contact. It is the average distance between the two objects from the contact boundary. (iv) **Containment** $f_c(A, B)$ is the percentage of volume of the supported object A lying within the 3D convex hull of B. This feature determines if an object is on another object or inside it. (v) A stable object has higher probability of supporting an unstable object. If the gravity line of an object is in alignment with the baseline, it is considered as stable. **Relative stability** $f_s(A, B)$ is defined as 1 if supporting object is stable and supported object is unstable, -1 for the reverse and 0 if both are stable or unstable.

3 Support Order Prediction



(a) (b)

Fig. 3. Support Order Prediction. (a)$\{O, O_1, O_2, O_3\}$: support in multiple hierarchy; $\{O_4, O_5, O_6\}$: containment, $\{O_7, O_8, O_9\}$: simultaneous support by multiple objects. (b)Reverse Level Order Traversal of Tree of Support. Edges in gray: Pruned edges. Nodes in light blue: skipped nodes.

The tree of support T is traversed using reverse level order traversal to predict the support order. The objects present at the leaf nodes are the ones not supporting any other object. So they are picked up first and then the upper layer is traversed and the process repeats until we reach the root node, i.e., the target object. Various special cases are identified and handled during tree traversal to ensure minimal damage during manipulation. In case of support by multiple hierarchy, the child node corresponding to the supported object is connected to multiple parent nodes from different layers. It is not feasible to retain all edges connecting to the child node. Retaining any of the edges in the upper layer(s) implies that the object corresponding to the child node will be searched even after its removal. On the other hand, pruning the edges in the lower layer(s) will increase the chance of damage since the presence of the supported object will be ignored while picking the parent node(s) in the lower layer(s). Therefore, the edge(s) between the child node and the parent node(s) at the lowest layer are retained while pruning off edges connected to parent node(s) in the upper layer(s). If any object contains other objects inside it instead of merely supporting it, then it is grasped directly without the need to remove the contained objects. So prior to retrieving any node during traversal, if the support type for a node is found to be "containment", then, this node is not retrieved. Fig. 3(b) graphically demonstrates the tree traversal and support order prediction for objects shown in Fig. 3(a). We traverse from the leaf nodes towards the root node. The support order is predicted as $O_3 \rightarrow O_9 \rightarrow O_2 \rightarrow O_8 \rightarrow O_7 \rightarrow O_4 \rightarrow O_1 \rightarrow O$.

4 Experiments and Results

We have collected a dataset consisting of 50 images, point clouds and depthmaps with different levels of clutter using Kinect and have manually created annotation for each image, and ground truth support relationship for each pair of regions in all images (Refer Fig.4). The raw depthmaps are smoothened using an adaptation of colorization method by Levin *et al.* [5]. A 5-stage hierarchical segmentation approach proposed by Arbelaez *et al.* [1] is used for segmenting the images.

Table 1. Accuracy: Structure class & Support Inference

Inference	Structure class Inference		Support Class Inference	
Type	Training	Testing	Training	Testing
Ground Truth Regions	100	97.02	73.42	64.72
Segmented Regions	97.79	83.88	53.00	49.17

The output of the structure classifier affects support inference, since only the regions predicted as graspable objects are given as input to the support classifier. The support relationship of the regions missed by structure classifier can not be established. And, if a region is falsely classified as graspable object, we may end up inferring an infeasible support relation. Hence, it is important to achieve high accuracy in structure class prediction. The accuracy of structure

Fig. 4. RGBD dataset for Support Inference: the dataset consists of images with different types of support, heavy clutter and occlusion

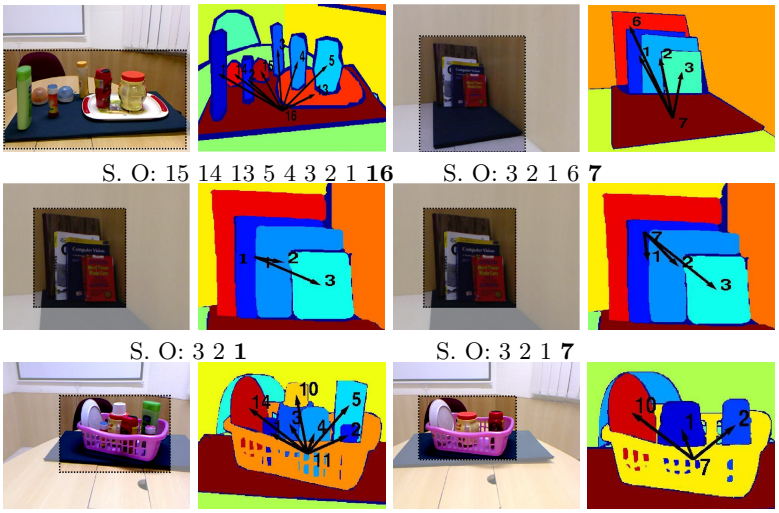


Fig. 5. Results of inference: The highlighted section in input images in column (a) & (c) are zoomed in columns(b) & (d) respectively for better view. The arrows point from target object to objects directly and indirectly supported by it. Support Order (S.O.) for each result is shown below the respective images.

class classification is shown in Table 1. High accuracy in structure class inference can be due to similar background entities in our dataset.

The result of support inference is shown in Fig. 5. The predicted support order corresponding to the images are shown below each image. The results in row 1, 2 and 3 depict support from below, support from side and containment respectively. The results in row 1 show support relation from below to multiple objects. In row 2, we can observe the hierarchical support relationship. Book 7 supports book 1 directly and book 1 supports books 2 and 3. Therefore, book 7 also indirectly supports books 2 and 3. The results in row 3 show containment. All the objects contained in the basket are shown as supported by the basket. But due to containment, they need not be removed in order to remove the basket, as evident in the support order shown for images in row 3.

The accuracy of support class classification is given in Table 1. We achieve 64.72% accuracy on the ground truth regions. We observe that, support inference

fails in a few situations. Often in frontal view, when the entire surface area of the supporting object is not visible and the contact to the supporting surface is not visible, the support relation is not inferred. In a few cases, due to error in structure class prediction, some valid graspable object regions are missed whereas some furniture regions are misclassified as graspable objects causing incorrect support inference. Sometimes, different planes of one object are segmented into different regions giving multiple regions for one object. In these cases also, infeasible support relations are inferred. Inaccurate segmentation errors also affect the support inference and thus support order prediction. As evident in Table 1, the accuracy of support class classification using segmented regions is 49.17%, which is significantly lower than compared to ground truth regions. This leaves lot of scope for improvement in this direction in future.

5 Conclusions

In this paper, we learned semantic interaction among objects in clutter by inferring support relationships and using them to predict the support order in which surrounding objects should be removed to access the target object. Our work extends the scope of semantic interaction to complex situations involving overlap, physical contact and objects of varied shape and size. We created a dataset consisting of different objects used in household and office environment and performed our experimentation on the same. In future, we plan to work on improving the performance of segmentation on RGBD data so that support relations are learned more accurately.

References

1. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: Contour detection and hierarchical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(5), 898–916 (2011)
2. Dogar, M., Hsiao, K., Ciocarlie, M., Srinivasa, S.: Physics-based grasp planning through clutter. In: RSS VIII (July 2012)
3. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, Manchester, UK (1988)
4. Hoiem, D., Efros, A.A., Hebert, M.: Recovering occlusion boundaries from an image. International Journal of Computer Vision 91(3), 328–346 (2011)
5. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Transactions on Graphics (TOG) 23(3), 689–694 (2004)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
7. Panda, S., Abdul Hafez, A.H., Jawahar, C.V.: Learning support order for manipulation in clutter. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS (2013)
8. Rasolzadeh, B., Björkman, M., Huebner, K., Kragic, D.: An active vision system for detecting, fixating and manipulating objects in the real world. The International Journal of Robotics Research 29(2-3), 133–154 (2010)

9. Rosman, B., Ramamoorthy, S.: Learning spatial relationships between objects. The International Journal of Robotics Research 30(11), 1328–1342 (2011)
10. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 746–760. Springer, Heidelberg (2012)
11. Sjoo, K., Jensfelt, P.: Learning spatial relations from functional simulation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1513–1519. IEEE (2011)
12. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008), `http://www.vlfeat.org/`

New Additive Wavelet Image Fusion Algorithm for Satellite Images

B. Sathya Bama*, S.G. Siva Sankari, R. Evangeline Jenita Kamalam, and P. Santhosh Kumar

Thigarajar College of Engineering, Department of ECE, Madurai, Tamil Nadu, India
{sbece,sivasankari,evangelinejenita}@tce.edu rsanfun@gmail.com

Abstract. Fusion of Low Resolution Multi Spectral (LRM) image and High Resolution Panchromatic (HRP) image is a very important topic in the field of remote sensing. This paper represents an efficient method for image fusion using New Additive Wavelet transform (NAW) based on the à trous algorithm. The fused image should preserve both geometric and the radiometric information. The use of geometric features along with spectral information improves the visualization quality of imagery. In this method intensity of the LRM image is added to the difference of the wavelet planes of HRP and LRP. The experimental results show that this method can well preserve spectral and spatial details of the source images. The proposed method provides competitive or even superior results for the input images compared to other well-known methods by providing 85.25% Quality with No Reference (QNR).

Keywords: correlation coefficient, discrepancy, MSSIM, radiometric distortion index, sobel grades, SSIM.

1 Introduction

Recent advancements in the field of remote sensing explore the opportunity of using images of different sensors for the same application. Pan sharpening is a kind of satellite image fusion method which injects the high-frequency components taken from the Pan image into resampled versions of the MS data to get a color image with pleasing and sharp appearance. Remote sensing calls for advanced pan sharpening methods to translate the spectral information of the coarse-scale MS data to the fine scale Pan image with minimal introduction of spectral distortions. Various methods have been proposed for pan-sharpening, which includes Intensity Hue Saturation (IHS) [1], the Principal Component Analysis and Multiresolution Transforms. The existing wavelet-based image fusion methods are mostly based on two computation algorithms: the Mallat algorithm and the à trous algorithm [2]. The paper concentrates on the à trous-algorithm.

Generally, the à trous-algorithm-based image fusion can be performed in two ways: 1) Substitute-Wavelet (SW) method and 2) Additive-Wavelet (AW) method [3]. The

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 313–318, 2013.

AW method maintains the high frequency of the HRP image and low frequency of the LRM image and the SW method eliminates the low frequency of the LRM image. However, if the same high frequency is injected into every LRM image, the AW method can produce redundant high frequency. As the HRP and LRM images generally have different local radiometry, radiometric distortion can arise in fused images. Also, the SW method can lose some of the information of the LRM image during substitution process [4].These problems can generate artifacts in fused image and it will suffer from radiometric and geometric distortions. To cope up with the above problem, this paper proposes a New Additive Wavelet algorithm based on à trous method. This paper is organized as follows section 2 describes the proposed method; section 3 discusses about the experimental results and section 4 conclude this paper.

2 New Additive Wavelet Transform

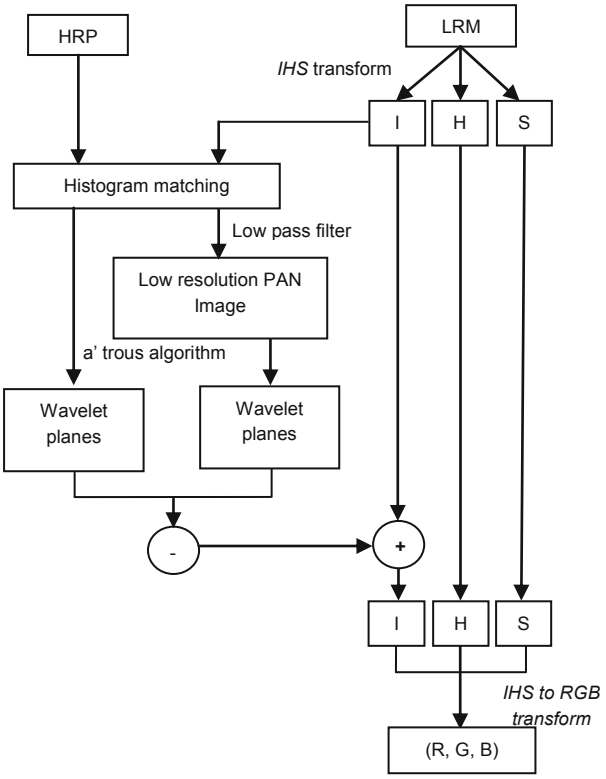


Fig. 1. New Additive Wavelet Transform

Fig 1 shows the proposed à trous based fusion method to decompose the image into wavelet planes .The New Additive Wavelet transform can be formulated by the following procedure.

1. Transform the RGB images into the IHS color space.
2. Perform histogram matching between the HRP image and the intensity.
3. Decompose the matched HRP image to n wavelet planes.

$$HRP = \sum_{l=1}^{n} w HRP_l + HRP^r \tag{1}$$

For an image a sequence of approximated images are obtained by convoluting with the filter that has B3 cubic spline profile. It is followed by the computation of the wavelet planes by the difference between two consecutive approximations.

4. Decompose the LRP image to n wavelet planes.

$$LRP = \sum_{l=1}^{n} w LRP_l + LRP^r \tag{2}$$

The LRP image is spatially degraded by filtering with a Gaussian low-pass filter.

5. Intensity of LRM image is added with the difference of wavelet planes of HRP and LRP images.

$$I_{new} = I + \sum_{l=1}^{n} w HRP_l - \sum_{l=1}^{n} w LRP_l \tag{3}$$

6. Transforming the new intensity together with the hue and saturation components back into RGB (inverse IHS transform).

   The proposed method considers a bandpass injection of high frequency extracted from an HRP image instead of conventional high-pass injection. Subsequently, the difference between the high frequency of the HRP image scale and the low frequency of the LRP image scale is injected. Thus, the low frequency of the LRP image scale is discarded. This implies that the high frequency between the LRM and HRP images is injected while maintaining the LRM image. Adding the wavelet planes to the intensity component preserves the spectral quantity more than that adding to the RGB of the LRM image individually. Since spectral values are present only in Hue and Saturation component, adding these planes to the intensity component does not affect the spectral values.

3 Data Sets and Experimental Results

Experiments have been conducted to evaluate the performance of the proposed method using the WORLDVIEW-2 data of Madurai, Tamil Nadu, India. For all data sets, the LRM image size is 128x128, and the HRP image size is 512x512 pixels. The

experiment area constitutes trees, buildings, and roads. The NAW method is compared with the existing methods.Fig.2 shows the HRP image of resolution 0.46m. Fig.3 is 1.84m LRM image interpolated at 0.46m.Fig 4 shows AWLP fusion method. The image obtained by NAW method is shown in Fig.5. This method produces high-quality image, implying that this image contains the necessary geometric information as well as rich radiometric information.

3.1 Quantitative Analysis

In Table 1, the reconstructed HR multispectral images are compared with the results produced by the existing methods such as IHS, wavelet, FFT enhanced IHS, Additive Wavelet (AW), Substitute Wavelet (SW). Table 1 demonstrates that the NAW the

Fig. 2. HRP image

Fig. 3. LRM image

Fig. 4. AWLP fused image

Fig. 5. NAW fused image

NAW method provides a less-distorted fused image compared with existing methods by finding Quality with No Reference (QNR) index [5]. In Table 2 shows the result of the spectral quality of the fused image measured by metrics such as discrepancy and Mean Structural-Similarity-Based Image Quality Assessment (MSSIM) [6].Spatial quality of the fused image is measured by metrics such as correlation coefficient [7]

and sobel grade. NAW yields the smallest spectral difference i.e. discrepancy and highest MSSIM i.e. more similar spectral values of LRM image. From Table 1-2 it's known that NAW gives superior performance for most of the quality metrics.

Table 1. Quality Indexes Of The Fused Images

Parameter / Methods	Radiometric distortion	Geometric distortion	QNR
IHS	0.0331	0.1183	0.8525
FFT Enhanced IHS	0.0141	0.1417	0.8467
Wavelet	0.0231	0.1341	0.8458
AWLP	0.0549	0.1220	0.8297
SW	0.0326	0.1207	0.8506
NAW	0.0309	0.1203	0.8525
Ideal	0	0	1

Table 2. Spectral and spatial quality Indexes Of The Fused Images

Parameter	Spectral Quality						Spatial Quality					
	Discrepancy			MSSIM			Correlation Coefficient			Sobel Grade		
Methods	R	G	R	G	R	G	R	G	B	R	G	B
IHS	11.96	17.66	16.65	0.16	0.21	0.41	0.87	0.96	0.90	42.5	52.1	46.3
FFT-Enhanced IHS	15.10	21.71	21.71	0.14	0.15	0.25	0.84	0.92	0.89	40.6	60.6	53.7
Wavelet	13.43	17.43	15.43	0.17	0.20	0.34	0.89	0.96	0.91	40.1	40.8	37.9
AWLP	22.20	30.13	22.23	0.18	0.19	0.32	0.85	0.93	0.88	40.8	54.8	52.7
sw	16.65	16.65	16.65	0.17	0.22	0.40	0.88	0.92	0.92	42.0	53.0	50.5
NAW	11.83	15.70	14.04	0.18	0.25	0.42	0.89	0.97	0.92	48.6	60.6	55.7

4 Conclusion

In this paper, New Additive Wavelet fusion method is proposed for preserving both radiometric information and geometric information. In this method bandpass injection of frequency by finding the difference between LRP and HRP image is done. The

proposed method can be considered as an improvement of the SW and AW methods in the sense that the LRM image is not decomposed. The proposed method provides comparatively better results than other existing methods in terms of spectral and spatial quality metrics.

References

1. Choi, M.: A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter. IEEE Trans. Geosci. Remote Sens. 44(6), 1672–1682 (2006)
2. González-Audícana, M., Otazu, X., Fors, O., Seco, A.: Comparison between the Mallat's and the 'à trous' discrete wavelet transform based algorithms for the fusion of multispectral and panchromatic images. Int. J. Remote Sens. 26(3), 597–616 (2005)
3. Núnez, J., Otazu, X., Fors, O., Prades, A., Pala, V., Arbiol, R.: Multiresolution-based image fusion with additive wavelet decomposition. IEEE Trans. Geosci. Remote Sens. 37(3), 1204–1211 (1999)
4. Stathaki, T.: Image Fusion Algorithms and Applications, 1st edn., ch. 6. Elsevier, Amsterdam (2008)
5. Alparone, L., Aiazzi, B., Baronti, S., Garzelli, A., Nencini, F., Selva, M.: Multispectral and panchromatic data fusion assignment without reference. Photogramm. Eng. Remote Sens. 74(2), 193–200 (2008)
6. Wang, Z., Bovik, A.C., Sheikh, H.R.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Transactions on Image Processing 13(4) (April 2004)
7. Zhou, J., Civico, D.L., Silander, J.A.: A wavelet transform method to merge landsat TM and SPOT panchromatic data. International Journal of Remote Sensing 19(4) (1998)

Iris Recognition under Non-ideal Imaging Conditions and CCD Noise

P.V.L. Suvarchala[1], S. Srinivas Kumar[2], and B. Chandra Mohan[3]

[1] Research Scholar, JNTUK, Kakinada, A.P, India
suvarchala_pvl@yahoo.com
[2] Department of ECE, UCE, JNTUK, Kakinada, A.P, India
samay_ssk2@yahoo.com
[3] Department of ECE, Bapatla Engineering College, Bapatla, A.P, India
chandrabhuma@gmail.com

Abstract. Iris recognition is the most reliable and dependable biometric system as the features of human eye are invariant and distinctive for every individual. Present iris recognition algorithms are tested using the bench mark databases which are assumed to be almost ideal except for eyelid and eyelash occlusions and rotational inconsistencies. It has been discussed elaborately by Daugman in [3] that, non-ideal imaging conditions affect the "authentics" distribution in the decision environment graph. Getting motivation from this observation, all possible non-ideal imaging conditions and Charge Coupled Device (CCD) noise are simulated and applied on the available databases. Legendre moments, introduced by Teague can achieve translation and scale invariance and also, close to zero value of redundancy measure, so that the moments correspond to distinct and autonomous features of the image. In the proposed method it is proved that, they can also work very well on noise affected features when trained and tested using SVMs. The performance of the Exact Legendre Moments (ELM) on UBIRIS and CASIA datasets proves to be very good with Correct Recognition Rate (CRR) = 99.6% under non-ideal imaging conditions and CCD noise.

Keywords: Iris recognition, Exact Legendre Moments, SVMs, CCD noise, Non-ideal imaging.

1 Introduction

Iris recognition is well accepted by public even though the imaging procedure requires the subject's cooperation, because it is unobtrusive, safe and accurate. Iris is the annular segment amid dark pupil and white sclera in the human eye which is an internal organ yet externally visible and has an unusual texture that is exclusive for each individual [2]. The textural features like freckles, ridges, corona and collarette region are alike all over the life time and do not alter at all. Iris biometric is well acknowledged by the public as the imaging technique is non-invasive and it also gained the interest of researchers as the iris prints

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 319–326, 2013.

are stable and unique. Several international airports have set up the iris scan systems to recognize their passengers and facilitate quick processing.

In CASIA database [6], the iris image is captured using Near Infra Red (NIR) wavelength camera, which is almost ideal except for the eyelash and eyelid occlusions where as in UBIRIS database [9], the iris image is captured using visual wavelength imaging device. In WVU data set, eye image is acquired using NIR illumination but iris images undergo rotational discrepancy. Visible wavelength illumination is not harmful to eye, but the images suffer from noise terms like poor contrast, low brightness, specular reflections and problems like blinking of the eye due to involuntary reaction, contraction of pupil to adjust to the illumination etc. NIR illumination is not safe but the images have good contrast and no specular reflections. It is observed that all the iris datasets available suffer from one of the issues of imaging conditions mentioned above but not all. The later versions of CASIA and UBIRIS databases are obtained under non ideal conditions like poor contrast, low brightness, motion blur and eye lash and eye lid occlusions but not CCD noise, hence making them suitable for testing the robustness of the segmentation method.

It was illustrated with proofs that the quality of the imaging strongly influences the "authentics" distribution in decision environment graph by Daugman in [3]. The camera gain, tilt, illumination, contrast, CCD noise etc., affect the image quality. The imaging devices are CCDs. During acquisition and conversion of photons in to digital signals by CCD sensors, various noise components like Thermal noise, Read-out noise, Fixed pattern noise, photon Shot noise are added to the signal, all as a whole called as CCD noise. Phase codes were used rather than amplitude information in multi scale quadrature method in [3] to encounter the imaging quality issues.

Legendre moments were used for scale invariant and translation invariant textural feature extraction in [14]. Hence, by creating non ideal iris database in which the images suffer from all possible irregularities like, blur due to camera motion and defocus, low contrast and brightness and mainly CCD noise, the robustness of the new feature extraction method with ELMs for iris recognition is tested. In the method proposed the CCD noise, simulated using the model proposed in [10] at various levels of variance along with camera motion blur is added to the CASIA and UBIRIS iris databases. The noise corrupted iris region is segmented normalized and the textural features are extracted using ELMs. The coefficients of all noisy iris images of different classes are picked out and then given ranking based on their entropies. The optimal ranked feature vectors are then classified using Support Vector Machines (SVM).

The rest of the paper is as follows. The CCD noise simulation, segmentation, normalization, textural attribute collection and feature vector generation using ELMs is explained in section 2. Section 3 is dedicated to the explanation about classification using SVMs. In section 4 the experimental results, comparisons and discussions are illustrated and in section 5, the conclusions and scope of future work are furnished.

1.1 Related Work

A detailed survey on iris recognition algorithms is presented in [1]. Zero crossing method with dissimilarity functions for matching was employed by Boles. The iris textural data were extracted using 2D Haar transform by Lim. Multi channel Gabor filtering for the extraction of iris features was used in [5]. Hilbert transform for iris feature extraction was used by Tisse in his method. Levelsets were applied for segmentation and wavelet features were extracted in [7]. In the proposed method a new feature extraction concept using ELMs is proposed and evaluated on the simulated non ideal databases.

1.2 Proposed Algorithm

A series of steps to generate non-ideal imaging conditions, applying them on available iris datasets, extracting features from images and finally classifying the data presented by the algorithm is as follows:

- **Step 1:** First the iris image from the database is taken, scaled to 1024 levels, after which its irradiance image is generated using Inverse Camera Response Function (ICRF).
- **Step 2:** Simulation of non-ideal imaging conditions in MATLAB
  - The intensity independent and intensity dependent CCD Noise is generated using two random noise sources of three different combinations with normal distribution of standard deviations chosen to be $\sigma_{independent} =$ 6% and $\sigma_{dependent} = 5\%, 10\%$ & 16%. and applied on the image.
  - Camera motion blur and poor contrast are also simulated and applied on the image.
- **Step 3:** The intensity image is generated from the noisy and blurred irradiance image using Camera Response Function (CRF) and then it is rescaled to 256 levels. The resulting image is corrupted with simulated CCD noise and camera motion blur.
- **Step 4:** The pre-processing, segmentation and normalization of iris region are done to facilitate feature extraction.
- **Step 5:** Then feature extraction using ELMs is used to generate feature vectors.
- **Step 6:** Finally classification of feature vectors using SVMs is done to evaluate the performance of ELMs on CCD noise and camera motion blur affected features in terms of % CRR.

2 Segmentation and Feature Extraction

The noise model of a CCD camera is proposed in [10] is used to generate synthetic CCD noise.

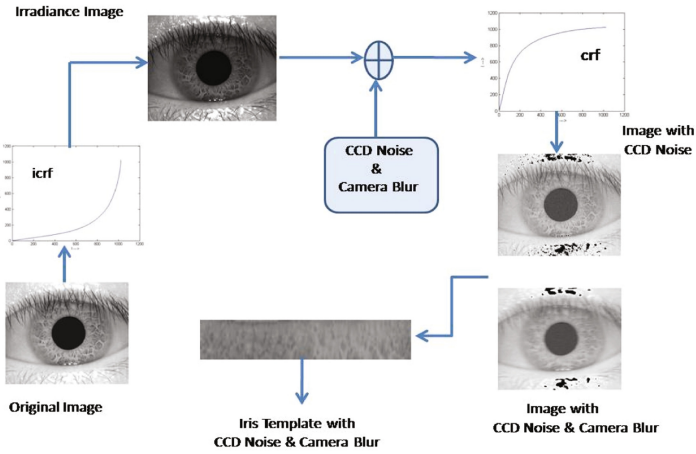


Fig. 1. CCD noise & blur simulation procedure

2.1 Iris Segmentation

In any biometric system, preprocessing of the image and extraction of distinct features from it, play vital role in deciding the recognition rate. The segmentation is done using Integro differential operator and normalization using the Rubber sheet model proposed in [2].

2.2 Feature Extraction with Exact Legendre Moments

Legendre moments generated with Legendre polynomials as kernel function, were first introduced in [13]. The 2D ELMs proposed by Hosny [12] are of order $(p+q)$ for an image intensity function $f(x, y)$, are defined as follows:

$$L_{pq} = \sum_{i=1}^{M} \sum_{j=1}^{N} I_p(x_i) I_q(y_j) f(x_i y_j) \tag{1}$$

where,

$$I_p(x_i) = \frac{(2p+1)}{(2p+2)} [xP_p(x) - P_{p-1}(x)]_{U_i}^{U_(i+1)} \tag{2}$$

and

$$I_q(y_j) = \frac{(2q+1)}{(2q+2)} [yP_q(y) - P_{q-1}(y)]_{V_j}^{V_(j+1)} \tag{3}$$

In the proposed method, ELMs are extracted from noisy and blurred iris images, the computation of which are fast and accurate.

3 Iris Classification with SVMs

In Iris recognition several researchers applied Hamming distance for pattern matching [1]. Conventional SVMs and Non symmetrical SVMs were applied by some researchers to differentiate the False Positive and False Negative cases and asymmetrical adaptive SVMs were used in [7] to reduce the matching times of the test samples.

SVMs have been an excellent tool for data classification. The fundamental idea is to map the data points in to a high dimensional space and separating them by a hyper plane with maximal margin. In the proposed method the iris feature vectors made up of ELMs are first ranked based on their entropies and then trained and tested using traditional SVMs with Linear kernel.

ELMs have least redundancy measure. Also ranking is given to the features according to their entropies. The entropy of the configuration which is in an order is lower than the entropy of the configuration with disorder. Therefore the irrelevant feature vector has higher entropy than the relevant feature vector. The feature with least entropy is given first rank and so on as proposed in [8], The entropy is calculated as follows. Entropy,

$$E = -\sum_{i=1}^{N}\sum_{j=1}^{N} d_{i,j} \times \log(d_{i,j}) + (1 - d_{i,j}) \times \log(1 - d_{i,j}) \tag{4}$$

where $d_{i,j}$ is the distance between features of N different samples in the same class. The features that have least entropy are more significant for classification.

3.1 Classification Evaluation

The classification of iris feature vectors is carried out using OSUSVM package available at `http://kaz.dl.sourceforge.net/project/svm/svm/3.00/osu-svm-3.0.zip`. Once the features are ranked according to their entropies, the class label is appended to each feature vector and the samples are classified based on supervised learning method. The proposed method is checked using 3-fold cross validation also, with 70% of the sample data used as train set and 30% of the data set as the test set. The multi class SVMs operated in One-against-All approach and are tuned with linear support vector classifier kernel. The $\%CRR$ is defined as

$$\%CRR = \frac{Correctly\ Recognized\ Users\ Number}{Total\ Number\ of\ Users\ Enrolled} \times\ 100 \tag{5}$$

4 Experimental Results

The experiments are conducted on a PC with Pentium i3, 2GHz processor and 2 GB RAM in MATLAB 7.10 environment. The evaluation of the proposed method is carried out on CASIA and UBIRIS databases under different non-ideal conditions. CASIA version-1 database consists of 756 iris images taken

in two sessions from 108 subjects. Each iris image is of 8-bit gray scale and resolution of 320 × 280. The UBIRIS version-1 database contributes 1205 iris images from 241 subjects with 5 samples from each subject with resolution of 150 × 200.

4.1 Discussion on Results

The performance of Legendre moments is already proved in terms of scale invariance and translation invariance in [14]. It is empirically proved by the results that the ELMs can work very well on CCD noise corrupted and blurred images. Experiments are carried out under three sets of different non ideal conditions, high CCD noise and camera blur ($\sigma_{dependent} = 5\% + \sigma_{independent} = 4\%$), medium CCD noise and blur ($\sigma_{dependent} = 10\% + \sigma_{independent} = 4\%$), and low CCD noise and blur($\sigma_{dependent} = 16\% + \sigma_{independent} = 6\%$) and motion blur of length 9 pixels and angle 5^o. Before extraction of ELM features, preprocessing step is

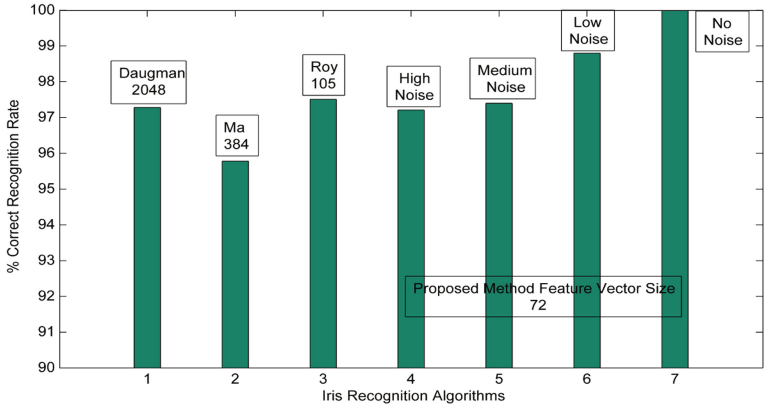


Fig. 2. Results of iris recognition algorithms on UBIRIS database: A comparison of % CRR and feature set size. The proposer's name, and feature set size is given in the boxes.

performed to enhance the contrast and highlight the blurred edges in the image by applying high boost filtering. Then iris region is segmented and normalized to get a 20 × 240 size noisy and blurred template. It is made into two blocks of size 20 × 120 for which ELMs of level-5 are extracted and resulting coefficients are arranged into a row forming a feature vector. The number of coefficients generated depends on the degree of Legendre polynomial. The higher the degree of polynomial, the more the number of coefficients and the better the performance of the proposed method. As there is a constraint regarding the feature set size, the degree of the polynomial chosen is 5, and the resulting feature vector length is 72. The performance of the proposed method is investigated in terms of % CRR and feature set size .

The classification of iris images is carried under different train:test ratios such as 6:1, 5:2, 4:3, 3:4, 2:5 and 1:6 in CASIA dataset and 4:1, 3:2, 2:3 and 1:4 in UBIRIS. The results on both datasets are shown in figures 3 and 4. It is found that the CRR is 100% for all train versus test ratios in UBIRIS and CASIA except for 1:6, under ideal, no noise conditions. It is also observed that the % CRR is slightly decreased as the number of samples for training is decreased as against the number of samples for testing. Another observation is that the % CRR is very good when compared to other methods even in high noise and blurred environment which proves the robustness of ELMs against CCD noise and blur and the number of features is also very small when compared with other methods which is shown in figure 2. The % CRR is $\geq$ 99.6% for UBIRIS and $\geq$ 99% for CASIA in the 3-fold cross validation test under all sets of CCD noise and blur conditions.

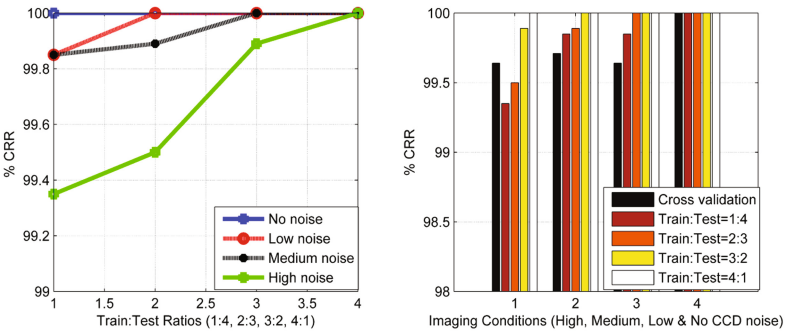


Fig. 3. Result of UBIRIS database for various non ideal conditions

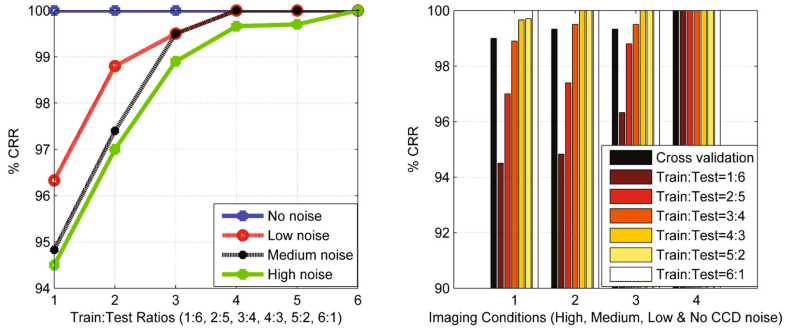


Fig. 4. Result of CASIA database for various non ideal conditions

5 Conclusions

It is observed from the experiments that, even in presence of high levels of noise also the % CRR is very high and it can still be improved with perfect segmentation. The ranked ELM feature set size is very small (72) when compared to other methods. In future, feature extraction using ELMs will be evaluated on higher versions of UBIRIS and CASIA databases.

Acknowledgement. The UBIRIS version-1, data set is owned by the department of computer science, University of Beira Interior, Portugal and the CASIA version-1, data set is owned by Institute of Automation, Chinese Academy of Sciences, China.

References

1. Bowyer, K.W., Hollingsworth, K., Flynn, P.J.: Image Understanding for Iris Biometrics: A survey, Comput. Vis. Image Underst. 10, 281–307 (2008)
2. Daugman, J.G.: How Iris Recognition Works. IEEE Transactions on Circuits and Systems, VideoTechnology 14(1) (2004)
3. Daugman, J.G.: New Methods in Iris Recognition. IEEE Trans. Syst. Man and Cyber. Part-B 37, 1167–1175 (2007)
4. Lim, S., Lee, K., Byeon, O., Kim, T.: Efficient Iris Recognition through Improvement of Feature Vector and Classifier. ETRI J. 23(2), 61–70 (2001)
5. Ma, L., Wang, Y., Tan, T.: Iris Recognition Based on Multichannel Gabor Filtering. In: Proc. Fifth Asian Conf. Computer Vision, vol. I, pp. 279–283 (2002)
6. Institute of Automation, Chinese Academy of Sciences, CASIA irisdatabase, `http://nlprweb.ia.ac.cn/english/irds/irisdatabase.htm`
7. Roy, K., Bhattacharya, P., Suen, C.Y.: Towards Nonideal Iris Recognition Based on Levelset Method, Genetic Algorithms and Adaptive Asymmetrical SVMs. Engineering Applications of Artificial Intelligence 24, 458–475 (2011)
8. Tan, F., Fu, X., Zhang, Y., Bourgeois, A.G.: Improving Feature Subset Selection using a Genetic algorithm for Microarray Gene Expression Data. IEEE Cong. Evol. Comp., 2529–2534 (2006)
9. UBIRIS dataset obtained from Department of Computer Science. University of Beira Interior, Portugal, `http://iris.di.ubi.pt`
10. Ce, L., Szeliski, R., Kang, S.B.: Automatic Estimation and Removal of Noise from a Single Image. IEEE Trans. Pattern Analysis and Machine Intelligence 30(2) (2008)
11. Grossberg, D., Nayar, S.K.: Modeling the Space of Camera Response Functions. IEEE Trans. Pattern Analysis and Machine Intelligence 26(10), 1272–1282 (2004)
12. Hosny, K.M.: Exact Legendre Moment Computation for Gray Level Images. Pattern Recognition 40, 3597–3605 (2007)
13. Teague, M.R.: Image Analysis via the General Theory of Moments. J. Opt. Soc. Am. 70(8), 920–930 (1980)
14. Chong, C.W., Paramesran, R., Mukundan, R.: Translation and Scale Invariance of Legendre Moments. Pattern Recognition 37(1), 119–129 (2004)

KID: Kirsch Directional Features Based Image Descriptor

B.H. Shekar[1], K. Raghurama Holla[1], and M. Sharmila Kumari[2]

[1] Department of Computer Science, Mangalore University, Karnataka, India
{bhshekar,raghu247}@gmail.com
[2] Department of Computer Science and Engineering, P.A. College of Engineering, Mangalore, Karnataka, India
sharmilabp@gmail.com

Abstract. In these days we have seen the development of local image descriptors for several computer vision applications in order to perform reliable matching and recognition. In this direction, we have made an attempt to propose a new local descriptor which uses the Kirsch's four directional edge features to describe the neighbourhood of the interest point. The descriptor computation mainly consists of two stages: feature detection (identification of interest points) and feature description. In the first stage, the interest points are detected using *Features from Accelerated Segment Test* (FAST) algorithm where interest points are identified by comparing the pixels on a circle of fixed radius around the interest point. In the second stage, the directional features for horizontal, vertical, right-diagonal and left-diagonal directions are extracted from the local region around the interest point using Kirsch masks. In order to achieve rotation invariance, the descriptor window coordinates are rotated with respect to the dominant orientation of the interest point. Experiments have been conducted on several image datasets to reveal the suitability of the proposed approach for various image processing applications. A comparative analysis with the other well known descriptors such as SIFT, SURF and ORB is also provided to exhibit the performance of the proposed local image descriptor.

Keywords: Interest point, feature detection, local descriptor, object recognition.

1 Introduction

The task of matching the local regions between two images affected by different kinds of geometric or photometric transformations such as translation, rotation, change in viewpoint, poor illumination etc, is a challenging problem in computer vision. In recent years, several researchers have focussed their attention on developing a local descriptor which is robust against the above challenges. We can find that the local descriptor based matching has many applications such as object recognition, face recognition, image retrieval, object tracking, video retrieval, stereo correspondence problem etc. The local features provide a powerful

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 327–334, 2013.

way of image representation and allows to recognize the objects or scenes within an image without performing image segmentation (see [20]).

Generally, the descriptor computation begins with the feature detection phase, where invariant features are extracted. These features must be distinctive and stable. In the next phase, the feature description is performed. It is important to choose the appropriate descriptor to represent the local region in the neighbourhood of the interest point. This descriptor must be robust against noise, geometric and photometric deformations. Then the descriptors are matched between different images. Hence we are motivated to take up the work on designing a new local image descriptor which is simple to implement, robust and efficient also.

The rest of the paper is organized as follows. Section 2 presents the review of existing algorithms for feature detection and feature description. The proposed model is introduced in section 3. Experimental results and comparison with other models are presented in section 4 and conclusion is provided in section 5.

2 A Review of Related Works

An overview of the literature for interest point detectors and feature descriptors is given below.

2.1 Feature Detectors: A Review

We can see that several algorithms for interest point detection have been reported in the literature [20]. The popularly used Harris corner detector proposed by Harris and Stephens [9] is based on the eigenvalues of the second moment matrix. Lowe(2004) [13] suggested a method for interest point detection by selecting the local extrema of an image, filtered with differences of Gaussians. Bay et.al(2006) [6] proposed a faster method of obtaining the Hessian detector using integral images and box filters for approximating the derivatives.

Rosten and Drummond [16] [17] developed a method for extracting interest points from an image using *Features from Accelerated Segment Test*(FAST) algorithm, which has become popular in recent years. The keypoints are identified by comparing the pixels in a circle of 16 pixels around the current pixel against a threshold. As the algorithm had some weaknesses, a machine learning approach is used to overcome the drawbacks of the initially developed algorithm. Finally, a non maximal suppression is performed to remove the multiple keypoints which are adjacent to one another. This detector is found to be more efficient when compared to DoG.

2.2 Local Descriptors: A Review

Several algorithms for computing the descriptors are available in the literature. A detailed comparative analysis can be seen in [14] [12] [8]. Lowe(2004) [13] introduced SIFT descriptor which consists of four stages, namely, scale-space peak

selection, keypoint localization, orientation assignment and descriptor building, resulting in a 128-dimensional descriptor. Ke and Sukthankar(2004) [10] developed PCA SIFT to improve the matching performance of SIFT. An extension of SIFT called *Gradient location-orientation histogram* (GLOH) [14] is proposed to increase the robustness and distinctiveness of SIFT, by computing the descriptor in log-polar location grid. Bay et al.(2008) [6] proposed SURF descriptor, which is computed from the sum of the Haar wavelet responses for x and y directions in the neighborhood of each keypoint.

To improve the computational efficiency of the high dimensional descriptors, binarization technique is suggested [7], where the descriptor vectors are mapped to Hamming space. The BRIEF descriptor is developed based on intensity difference tests. The binarized descriptors are compared using Hamming distance metric. Rublee et al. [18] developed the Oriented Fast and Rotated BRIEF(ORB), which is invariant to rotation and robust to noise. The ORB is an extension of BRIEF descriptor and the keypoints are found using FAST algorithm. Strecha et al. [19] suggested a method based on similarity-sensitive hashing algorithms applied to local SIFT descriptors to improve the performance of feature matching in wide-baseline stereo problems. Alahi et al. [5] proposed a descriptor called *Fast Retina Keypoint*(FREAK) motivated by the way in which the images are processed in the human visual system. It computes binary strings by comparing image intensities over a retinal sampling pattern.

We have seen that there are several techniques available for developing a local descriptor. However, the robustness to various challenges such as geometric and photometric transformations could not be achieved completely. So, we propose an efficient method of designing a local descriptor based on Kirsch's directional features to address the above challenges. The proposed model shows better results for several object datasets and face datasets, and hence suitable for real-time feature matching applications. The details of the proposed model are given below.

3 Proposed Model

In the proposed descriptor, the keypoints are extracted using the well known FAST detector as it is proved to be very efficient and robust in real time feature detection.

In order to achieve rotation invariance, the dominant orientation for each keypoint needs to be calculated. To obtain the orientation, we adopt the same method as applied in SURF descriptor [6] since the convolution operation becomes faster with the use of integral image. The Haar wavelet responses for x and y directions within the circular neighbourhood having a fixed radius around each keypoint are calculated. The dominant orientation is calculated by obtaining the sum of all responses within a sliding window of size $\pi/3$. The x and y responses within the window are added to get a local orientation vector. The longest such vector gives the orientation for the keypoint.

To extract the descriptor for each keypoint, a square window of size 16×16 pixels centred on the keypoint is considered. This descriptor window coordinates are aligned in the direction of dominant orientation of the keypoint. To describe the features in this local region we use Kirsch's features [11] as it gives the directional features. Also, Kirsch edge detector accurately detects the four directional edges since all the eight neighbours are considered in the operation. The directional gradient proposed by Kirsch is given by [15]:

$$G(i,j) = \max_{k=0}^{7}\{|5P_k - 3Q_k|\} \tag{1}$$

where,

$$\begin{aligned} P_k &= A_k + A_{k+1} + A_{k+2} \\ Q_k &= A_{k+3} + A_{k+4} + A_{k+5} + A_{k+6} + A_{k+7} \end{aligned} \tag{2}$$

In Eqn(1), $G(i,j)$ represents the gradient of the pixel at (i,j). The subscripts of A are evaluated modulo 8, A_k denotes the 8-neighbours of the pixel at (i,j) as shown in fig 1, $i = 0, 1, ...7$.

A_0	A_1	A_2
A_7	(i,j)	A_3
A_6	A_5	A_4

Fig. 1. 8-neighbours of a pixel (i,j)

Now, the four directional features are calculated as follows:

$$\begin{aligned} G(i,j)_H &= \max\left(|5P_0 - 3Q_0|, |5P_4 - 3Q_4|\right) \\ G(i,j)_V &= \max\left(|5P_2 - 3Q_2|, |5P_6 - 3Q_6|\right) \\ G(i,j)_{RD} &= \max\left(|5P_1 - 3Q_1|, |5P_5 - 3Q_5|\right) \\ G(i,j)_{LD} &= \max\left(|5P_3 - 3Q_3|, |5P_7 - 3Q_7|\right) \end{aligned} \tag{3}$$

where $G(i,j)_H$, $G(i,j)_V$, $G(i,j)_{RD}$ and $G(i,j)_{LD}$ represent the directional edge features for horizontal, vertical, right-diagonal and left diagonal directions respectively. The filter masks to be used for calculating directional feature vectors are shown in fig 2.

In the above approach, the global characteristics of images are not considered. So, the normalized image intensities within the local region is used as global feature in addition to the local directional features. Each of the 16×16 features within the local region are compressed to 4×4 using bicubic interpolation technique where the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighbourhood and then the descriptor is normalized. Thus we have a descriptor of length $4 \times 4 \times 5 = 80$. The fig 3 summarizes the different stages of descriptor building.

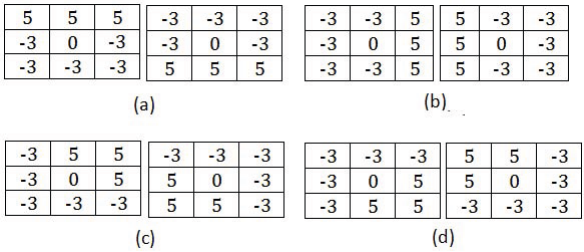


Fig. 2. Kirsch masks to be used for directional features, (a) horizontal, (b)vertical, (c)right-diagonal, (d)left-diagonal

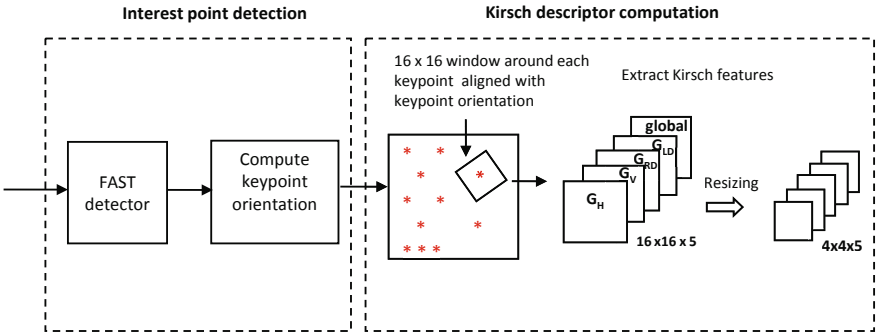


Fig. 3. Descriptor computation stages

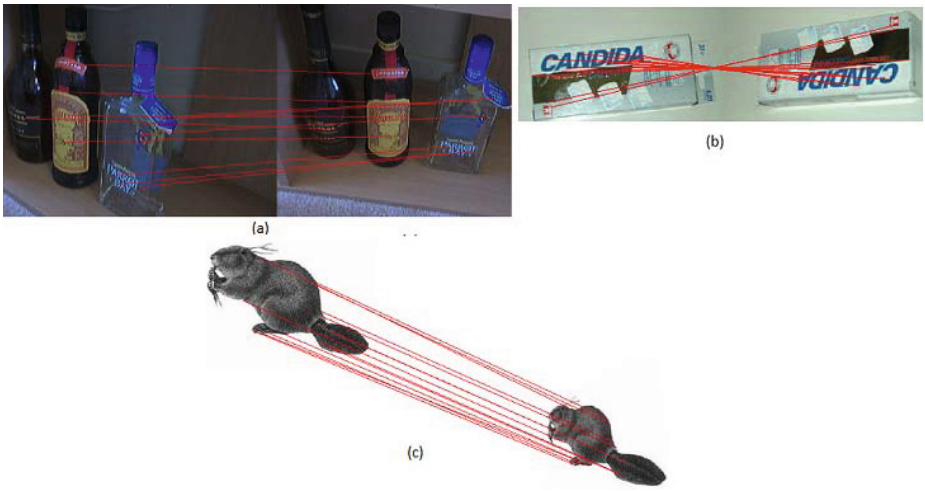


Fig. 4. Keypoints matching by Kirsch descriptor for different kinds of transformations

The descriptors are compared using City block distance measure. The nearest neighbour distance ratio(NNDR) [13] is used to match the keypoints. The two regions are matched by comparing the distance ratio between the first and second nearest neighbours against a threshold. If D_B and D_C are the first and second nearest neighbours of D_A, then the region is matched if $\frac{\|D_A - D_B\|}{\|D_A - D_C\|} < T$, we fixed the threshold value $T = 0.75$ empirically.

4 Experimental Results

In this section, we present the experimental results to reveal the performance of the proposed descriptor. We conducted experiments on several standard image datasets such as COIL-20 object dataset [2], 53-objects dataset [3], YALE face database [4], AT & T face database [1] etc. The experiments have been performed with various combinations of training and testing sets. Also, we have provided a comparative analysis with the other standard descriptors such as SIFT, SURF and ORB by performing the similar experiments for these descriptors on the same datasets. All the descriptors were tested with their default parameters settings(i.e. default threshold parameters for interest point detection and matching). The recognition rate is defined as the number of correctly recognized test objects divided by the total number of test objects. Some of the results are given in Table 1. The figure 4 demonstrates the result obtained by the proposed descriptor for matching the corresponding keypoints of two images affected by some transformations such as rotation, scale change, viewpoint change etc.

The COIL-20 dataset consists of gray scale images of 20 objects taken at a pose interval of 5 degree resulting in 72 images per each object, and it contains a total of 1440 images. In the first training set, we considered the images taken at a pose interval of 15 degree, resulting in 24 training images for each subject and the remaining images were used for testing purpose. The images taken at every 30 degree interval were chosen for the second training set, resulting in 12 training images for each object and the remaining 60 images were used for testing. In the third phase, the objects captured at a pose interval of 60 degree were selected for training, which leads to 6 training images for each subject and the remaining 66 images were used for testing purpose. The proposed model achieved the recognition rates: 98.75%, 96.67% and 86.97% respectively for the three different training sets. From Table 1, it can be noticed that the proposed descriptor outperforms the other descriptors on each test combinations. The SURF descriptor (with default parameters) produces weak results since the number of detected interest points are too less to be matched. The YALE face database consists of 165 face images with 15 subjects that include variation in both facial expression and lighting. In the first training set, the first four face images of each person were considered and the remaining images have been chosen for testing. The proposed model achieved a recognition rate of 95.23% for this test images. In the second phase, the first three face images of each subject were chosen for training and the remaining eight images were used for testing. The recognition

rate for this test set is 93.33%. Objects-53 dataset consists of images of 53 objects with five different views for each subject, containing a total of 265 images. The images are captured at different camera view points. The first, third and fifth images of each subject were chosen for the training set and the remaining images were selected for testing purpose. ORB descriptor produced good results for this dataset, followed by the proposed model, achieving a recognition rate of 82.69%. The AT & T face database contains 10 different images of 40 subjects. Some images were captured at different times. The dataset contains images with changes in lighting and facial expressions. In our experiment, the odd numbered samples were used for training set and the remaining images were used for testing purpose. The SIFT produced better accuracy, followed by the proposed model (98%). From the experimental results it can be noticed that the proposed

Table 1. Descriptor matching performance on different datasets

Dataset + (No. of TrainSamples per sub.)	Recognition rate(%)			
	proposed model	SIFT	SURF	ORB
COIL-20(24)	98.75	96.56	83.33	95.73
COIL-20(12)	96.67	93.42	72.00	92.25
COIL-20(6)	86.97	84.47	48.94	81.81
YALE(4)	95.23	95.23	85.71	79.05
YALE(3)	93.33	93.33	83.33	71.66
53-objects(3)	82.69	77.88	74.04	86.53
AT & T(5)	98.00	99.50	92.50	86.50
Average Results(%)	**93.09**	**91.48**	**77.12**	**84.79**

descriptor gives consistently better accuracy for all the databases tested when compared to other local descriptors. Other approaches produced better results for object datasets, however, the performance decreased when tested on face databases.

5 Conclusion

We have presented a robust and efficient local descriptor based on Kirsch's directional features for representing the texture feature of images. The interest point detection is carried out efficiently using FAST detector. The experimental results show that the proposed local descriptor is on a par with the standard descriptors such as SIFT, SURF and ORB. The proposed method is computationally efficient, simple to implement and hence suitable for many computer vision and video processing applications.

References

1. AT & T face database, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
2. COIL-20 object dataset, http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
3. object dataset, http://www.vision.ee.ethz.ch/showroom/zubud/
4. YALE face database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html
5. Alahi, A., Ortiz, R., Vandergheynst, P.: FREAK: fast retina keypoint. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510–517 (June 2012)
6. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
7. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: Binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
8. Gauglitz, S., Höllerer, T., Turk, M.: Evaluation of interest point detectors and feature descriptors for visual tracking. International Journal of Computer Vision 94, 335–360 (2011)
9. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, pp. 147–151 (1988)
10. Ke, Y., Sukthankar, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, pp. II-506–II-513 (2004)
11. Lee, S.W.: Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(6), 648–652 (1996)
12. Li, J., Allinson, N.M.: A comprehensive review of current local features for computer vision. Neurocomputing 71(10-12), 1771–1787 (2008)
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
15. Pratt, W.: Digital image processing. John Wiley & Sons (2007)
16. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
17. Rosten, E., Porter, R., Drummond, T.: Faster and better: A machine learning approach to corner detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(1), 105–119 (2010)
18. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 2564–2571 (November 2011)
19. Strecha, C., Bronstein, A., Bronstein, M., Fua, P.: LDAHash: Improved matching with smaller descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(1), 66–78 (2012)
20. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Found. Trends. Comput. Graph. Vis. 3(3), 177–280 (2008)

Local Morphological Pattern Spectrum Based Approach for Off-line Signature Verification

B.H. Shekar[1], R.K. Bharathi[2], and Bharathi Pilar[3]

[1] Department of Computer Science, Mangalore University, Karnataka, India
bhshekar@gmail.com
[2] Department of Master of Computer Applications, S J College of Engineering, Mysore, Karnataka, India
rkbharathi@hotmail.com
[3] Department of Master of Computer Applications, AIMIT, St. Aloysius College, Mangalore, Karnataka, India
bharathi.pilar@gmail.com

Abstract. In this paper, we present a local morphological pattern spectrum based approach for off-line signature verification. The proposed approach has three major phases : Preprocessing, Feature extraction and Classification. In the feature extraction phase, the signature image is partitioned into eight equally sized vertical blocks and local morphological pattern spectra of each block is obtained. The spectrum thus obtained for each block is converted to normalised ten bin histogram and to form a feature vector of the signature. The Earth Movers Distance (EMD) measure is used for classification and the performance is measured through FAR/FRR metric. Experiments have been conducted on standard signature datasets namely CEDAR and GPDS-160, and MUKOS, a regional language (Kannada) dataset. The comparative study is also provided with the well known approaches to exhibit the performance of the proposed approach.

Keywords: Pattern spectra, Earth Movers Distance, Histogram matching, Off-line signature verification.

1 Introduction

Over the last few decades, Biometrics has emerged as a convenient and reliable technology and has become one of the most active research areas due to the extensive potential applications in human−computer interaction and public security. The handwritten signature is recognized as one of the most widely accepted personal attribute for convenient certificate of consent and authorization. The handwritten signature verification can be performed automatically either on-line or off-line. On-line signature verification needs special instruments such as a tablet, stylus, or digitizer, where as off-line verification employs the static image of a signature. Off-line signatures are already popular as it does not require any special devices for registration and verification can also be done in the absence of the signer.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 335–342, 2013.

Off-line signatures are the static 2D image of the registered signature at a certain point of time. Processing the signatures imply, accessing the image in the absence of dynamic information which increases the complexity of verifying the test signature as genuine or classifying them from forge. In addition, there are many other inherent complications such as, the intraclass deviation of the signature samples, i.e variation of the signatures by the genuine signer due to age, illness, orientation of the document used to sign, pen width, deteriorated signatures, illegible signatures, and so on, which needs greater attention in signature verification. Some of the well accepted off-line signature verification approaches based on varying features, feature selection techniques and selection of classifiers are reviewed here.

Saburoni et al. [11] have proposed to compute pattern spectrum on local shape of signature. In this work, granulometric size distributions have been used to define local shape descriptors. The limitation of this approach is that it is similar to the human perception on local features where the minor components of the signature image is neglected and is computationally expensive. Pal et al. [8] have proposed a bi-script off-line signature identification considering both English and Bengali signatures. Solar et al. [10] concentrated on local interest points and descriptors for off-line signature verification. Kumar et al. [5] presents a novel set of local features based on surroundedness property of a signature image to provide a measure of texture through the correlation among signature pixels. Almazain et al. [1] calculates the pixel density distribution resulting in non-rigid feature extraction of the signature and demonstrated the performance on GPDS-100 and GPDS-750 datasets, but significant results are obtained for random forgeries only. To demonstrate that mere feature selection is not important in off-line signature verification, Shekar and Bharathi [12] concentrated on reducing the dimension of feature vectors, preserving the effective features obtained through eigen-signatures and later extended to Kernel eigen-signatures based on Kernel-PCA [13].

Even though ample number of algorithms are developed for off-line signature verification, devising an efficient off-line signature verification with high accuracy is still an open problem. Due to the inherent complexities of signature image, developing an accurate verification approach is much more complex. In this context, we proposed a simple and an accurate off-line signature verification approach that works well for any type of signature including different languages. The proposed approach is based on the newly introduced frame work called *Local morphological pattern spectrum*, which is suitable for off-line signature verification. We proposed to employ a EMD measure for the purpose of classification with suitable modifications. We have considered skilled forgery as a means to measure the accuracy (FAR/FRR). The intention behind considering skilled forgery is that the similarity of the skilled forge signature with the actual signature is very high when compared to random forged signature and hence classifying genuine signature from skilled forge signature accurately is highly complex. The proposed approach posses high accuracy when tested with skilled forged samples.

The remaining part of the paper is organised as follows. In Section 2, an insight on morphological patter spectrum is given. The proposed approach is given in section 3, followed by the experimental set-up along with the results and comparison with state-of-art approaches are brought out in section 4 and conclusion in section 5.

2 Morphological Pattern Spectrum Representation

Feature extraction is an important process not only in signature verification but also in any kind of pattern recognition problem. In this paper, unique structural features are extracted from the signature through the use of a novel method called morphological pattern spectrum. The pattern spectrum is a morphological tool that gives the quantitative information about the shape and sizes of the objects in the image. Spatial pattern spectrum provides the information on the distribution of the pixels in the binary image and are observed to be insensitive to the spatial information. A forward mask and its 90^o rotated backward mask is embossed to obtain the *distance transform map* which is similar to *skeleton strength map (SSM)* [6]. In this paper, we focus on spatial distribution of the signature points. The corresponding distance transform map value of each dominant pixel results in pattern spectrum of the signature.

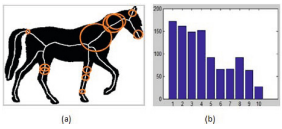


Fig. 1. (a) Maximal disc inscribed on the skeleton of the horse binary image (b) Histogram representing the radius of the disc and the number of such discs

The original pattern spectrum proposed in [7] is based on Serras Mathematical morphology filters that describe the distribution of local figure thickness. The thickness map is a $2D$ function, whose value in each point of the frame is equal to the maximum size of structural element, fully inscribed in the image, covering the point (the values are positive inside a shape) or on the background (the values of thickness are negative on the background). In case of a discrete set of sizes of the structuring elements, Maragos spectrum is just a histogram of values of the thickness map. In figure 1(a) skeleton of the horse binary image is taken. The radius of the maximal disc inscribed inside the shape for each skeleton point is computed. The number of such discs for each radius is calculated. A histogram representing the radius of the disc and the number of such discs is shown in the figure 1(b)).

3 Earth Movers Distance

The Earth Movers Distance (EMD) is defined as a minimal cost that must be paid to transform one distribution into the other [9]. It is more robust than histogram matching techniques and can operate on variable-length representations of the distribution. For instance, if Ha and Hb are two histograms, EMD is the minimum amount of work needed to transform histogram Ha towards Hb. Given two distributions, one can be seen as a mass of earth properly spread in space and the other as a collection of holes in the same space. Then, the EMD measures the least amount of work needed to fill the holes with earth.

To illustrate, let P = $(p_1, wp_1), ..., (p_m, wp_m)$ be the first distribution with m clusters, where p_i is the cluster representative and wp_i is the weight of the cluster and Q = $(q_1, wq_1), ..., (q_m, wq_m)$ be the second distribution with n clusters. In other words, $p_i$and q_i typically represent bins in a fixed partitioning of the relevant region of the underlying feature space, and the associated wp_i and wq_i are a measure of the mass of the distribution that falls into the corresponding bin. Let D= d_{ij} be the ground distance matrix where d_{ij} is the ground distance between clusters p_i and q_j. We need to estimate the flow F = $[f_{ij}]$, with f_{ij} be the flow between p_i and q_j , that minimizes the overall cost. The Minimal cost (WORK), between two distribution is calculated as follows and subjected to the constraints:

$$WORK(P,Q,F) = \sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij} \qquad (1) \sum_{j=1}^{n} f_{ij} \leq W_{pi}, 1 \leq i \leq m;$$

$$(2) \sum_{i=1}^{m} f_{ij} \leq W_{qi},\ 1 \leq j \leq n; \quad (3) \sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = min\left(\sum_{i=1}^{m} W_{pi}, \sum_{j=1}^{n} W_{qj}\right)$$

Constraint(1) specifies that the supplies can be moved from P to Q and not vice versa. Constraint(2) limits the amount of supplies that can be sent by the clusters in P to their weights. Constraint(3) limits the clusters in Q to receive no more supplies than their weights; and constraint(4) forces to move the maximum amount of supplies possible. The amount of flow is called as the total flow.Once the transportation problem is solved, and the optimal flow F is found, the earth movers distance is defined as the resulting work normalized by the total flow:

$$EMD(P,Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}}$$

4 Proposed Approach

In the proposed work, the local morphological pattern spectrum is used to obtain the local histograms of the signature partitions. The histogram based features of all the partitions are accumulated and EMD distance measure is used to obtain the distance between the signatures. Based on the distance, the testing signature

is verified as genuine or forge. In the preprocessing phase, the given signature image is binarized using Otsu's binarization method. The noise intruded due to binarization is eliminated using morphological filter operations. A bounded box is fitted to the actual signature boundary region, thus eliminating the non signature portion of the image.

4.1 Feature Extraction

Each of the preprocessed signature image results in a matrix with the pixel intensity values either 0 or 1.The preprocessed signature image is vertically partitioned into 8 blocks and corresponding local morphological pattern spectrum is obtained for each block. Now, each preprocessed signature image in the dataset, being represented in a matrix form, is transformed into a local morphological pattern spectrum histogram, resulting in a feature vector. The overall algorithm to obtain a local morphological pattern spectrum is presented below:

Fig. 2. (a)Preprocessed image(b)8 partitions(c)Local morphological pattern spectra

(a) The binarized and preprocessed training signature image is enclosed in a bounding box (figure 2(a)) is partitioned into 8 equally sized vertical blocks (figure 2(b)).
(b) The radius of the maximal disc inscribable inside the rectangular block for each signature pixels in each partition is calculated. Here, the signature block boundary forms the reference line for the radius calculation.
(c) In order to obtain distance transform map, $3X3$ structuring element is used as forward and backward mask. Pattern spectrum is obtained from the distance transform map on each signature pixel of the signature image resulting in local morphological pattern spectrum.
(d) A local morphological pattern spectrum thus obtained from above steps is converted to normalised 10 bin histogram. Hence, for an input signature image with 8 partitions, $10*8$ dimensional feature vector is obtained.
(e) The above steps are repeated for all the training samples of all the signers to obtain *'local morphological pattern spectrum knowledge base'* considering both genuine and skilled forgery samples of the respective datasets.

4.2 Classification

Given a signature for its classification, the match is obtained by EMD distance measure. However as there are multiple pattern spectrum for every signature, the distance between any two signature say, $S_1$and S_2 is computed as follows:

- Let S_1 and S_2 are the two signatures to be compared.
- Let $H_i = \{H_{i1}, H_{i2}...H_{i8}\}$ and $H_j = \{H_{j1}, H_{j2}...H_{j8}\}$ are the pattern spectrum features of the signature S_1 and S_2 respectively.
- Now the local distance d_k of each partition is obtained using EMD : $d_k = EMD(H_{ik}, H_{jk})$, where k = 1,2,...8.
- The total EMD distance $D = \sum_{k=1}^{k=8} d_k$

The test signature sample is then classified as either genuine or forge and belong to some class based on the minimum distance D that we have obtained by comparing with all the training samples stored in the *local morphological pattern spectrum knowledge base*.

5 Experimental Results and Discussions

Experiment of the proposed approach is conducted on standard off-line English signature datasets namely: CEDAR (The Centre of Excellence for Document Analysis and Recognition) and GPDS-160,a sub-corpus of GPDS-300 (Digital Signal Processing Group (GPDS) of the Universidad de Las Palmas de Gran Canaria). CEDAR consists of a total of 2640 signatures obtained from 55 signers (24 genuine and 24 skilled froge), whereas in GPDS-160 contributed 8640 samples from 160 signers (24 genuine and 30 skilled forge). In addition, we have also extended the experiments on our regional language off-line signature corpus called MUKOS (Mangalore University Kannada Off-line Signature). In this corpus we have collected 30 genuine and 15 skilled forge signature samples from each of the 30 signers resulting in a total of 1350 samples. All experiments are conducted using MATLAB tool and tested on Pentium(R) dual core CPU with 3GB RAM on Windows-7.

The knowledge base contains the pattern spectrum of every signature in the data set including both genuine and skilled forge samples. For each dataset, the signature samples are divided into two groups: training sample set and testing sample set with varying number of samples. We have carried out four set of experiments. In *Set-1*, first 10 genuine and 10 random forge(RF) samples are considered as training samples. In *Set-2* we have taken first 15 genuine along with 15 random forge(RF) samples to train and tested against the remaining samples along with the skilled forge samples of the respective datasets. In *Set-3*, randomly chosen 10 genuine samples(Rnd) are considered for training, and tested with the remaining samples(Rmng), and in *Set-4*, 15 samples are chosen randomly from the respective datasets for training along with the random forgeries and remaining samples are considered for testing. In order to over come the effect of the randomness, Set-3 and Set-4 experimentations are repeated five times and the average result is tabulated. Here, random forgeries are the genuine samples of the other signers of the same datasets.

The metrics FAR and FRR obtained for CEDAR dataset along with a comparative analysis with the state-of-art approaches is tabulated in Table 1.

Table 1. Experimental Results obtained for CEDAR Dataset

Experimental Set-up	**Training No.**	**Testing No.**	**Accuracy**	**FRR**	**FAR**
Set-1	10G+10RF	14G+24SF	89.66	10.44	12.33
Set-2	15G+15RF	9G+24SF	**90.42**	**9.68**	**10.98**
Set-3	10G(Rnd)+10RF	14G(Rmng)+24SF	88.34	11.76	12.68
Set-4	15G(Rnd)+15RF	9G(Rmng)+24SF	89.80	10.20	11.04
Results obtained for CEDAR Dataset - A comparison:					
Proposed by	**Feature Type**	**Classifier**	**Accuracy**	**FRR**	**FAR**
Kalera et al. [4]	Word Shape	PDF	78.50	22.45	19.50
Chen and Shrihari [2]	Zernike moments	DTW	83.60	16.60	16.30
Kumar et al. [5]	Signature Morphology	SVM	88.41	11.59	11.59

Table 2. Experimental Results obtained for GPDS-160 Dataset

Experimental Set-up	**Training No.**	**Testing No.**	**Accuracy**	**FRR**	**FAR**
Set-1	10G+10RF	14G+30SF	86.56	13.54	14.64
Set-2	15G+15RF	9G+30SF	89.18	10.82	11.64
Set-3	10G(Rnd)+10RF	9G(Rmng)+30SF	87.13	12.87	11.77
Set-4	15G(Rnd)+15RF	9G (Rmng)+30SF	**91.06**	**9.40**	**10.63**
Result obtained for GPDS-300/160 dataset : A comparative analysis					
Model Proposed	**Feature type**	**Classifier type**	**Accuracy**	**FRR**	**FAR**
Ferrar et al. [3]	Geometric features	SVM	86.65	15.41	13.12
		HMM		14.10	12.60
Vargas et at. [14]	GLCM	SVM +LBP	87.28	22.49	6.17
Solar et al., [10]	Local interest points	Bayseian	84.70	16.40	14.20

Table 3. Experimental Results obtained for MUKOS Dataset

Experimental Set-up	**Training No.**	**Testing No.**	**Accuracy**	**FRR**	**FAR**
Set-1	10G+10RF	20G + 15SF	93.36	6.64	10.69
Set-2	15G+10RF	15G + 15SF	93.26	6.78	7.58
Set-3	10G(Rnd)+10RF	20G(Rmng) + 15SF	**94.02**	**5.98**	**7.35**
Set-4	15G(Rnd)+10RF	15G(Rmng) + 15SF	93.76	6.24	7.63
Results for MUKOS dataset- A comparative analysis					
Proposed by	**Method**	**Classifier**	**Accuracy**	**FRR**	**FAR**
Shekar and Bharathi [12]	eigen signature	Euclidean distance	93.00	6.40	11.07

Table 2 gives the result obtained for GPDS-160 dataset on the proposed approach with the comparative analysis. In Table 3, the performance of the approach on MUKOS is tabulated along with the results of our earlier approach.

6 Conclusion

In this paper, we explored the application of morphological pattern spectra on off-line signatures to extract the local pattern spectra based features followed by classification using EMD. The input image is pre-processed and partitioned into 8 vertical blocks. The pattern spectrum is obtained for each block and the corresponding local pattern spectrum for each block is computed followed by EMD

based distance computation for classification purpose. Extensive experimentation is conducted on well known publicly available signature dataset :CEDAR and GPDS-160 and a regional language signature dataset called MUKOS. In order to highlight the superiority of the proposed approach, a comparative analysis is provided with the state-of-the-art off-line signature methods. It is found that the proposed approach is simple to implement, computationally efficient and accurate in terms of classification.

Acknowledgements. The authors would like to thank DST, Govt. of India,for supporting this work under the project: INT/RFBR/P-133.

References

1. Almazan, J., Fornes, A., Valveny, E.: A non-rigid feature extraction method for shape recognition. In: ICDAR, pp. 987–991. IEEE (2011)
2. Chen, S., Srihari, S.: Use of exterior contours and shape features in off-line signature verification. In: ICDAR, pp. 1280–1284 (2005)
3. Ferrer, M., Alonso, J., Travieso, C.: Offline geometric parameters for automatic signature verification using fixed-point arithmetic. IEEE-PAMI 27(6), 993–997 (2005)
4. Kalera, M.K., Srihari, S., Xu, A.: Off-line signature verification and identification using distance statistics. IJPRAI 18, 228–232 (2004)
5. Kumar, R., Kundu, L., Chanda, B., Sharma, J.D.: A writer-independent off-line signature verification system based on signature morphology. In: ICIITM, pp. 261–265. ACM (2010)
6. Latecki, L.J., Li, Q.N., Bai, X., Liu, W.Y.: Skeletonization using ssm of the distance transform. In: IEEE-ICIP, vol. 5, pp. 349–352 (2007)
7. Maragos, P.: Pattern spectrum and multiscale shape representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7), 701–716 (1989)
8. Pal, S., Alireza, A., Pal, U., Blumenstein, M.: Off-line signature identification using background and foreground information. In: ICDICTA, pp. 672–677. IEEE (2011)
9. Rubner, Y., Tomasi, C., Guibas, L.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40(2), 99–121 (2000)
10. Ruiz-Del-Solar, J., Devia, C., Loncomilla, P., Concha, F.: Offline signature verification using local interest points and descriptors. In: CIARP 2008, pp. 22–29 (2008)
11. Sabourin, R., Genest, G., Preteux, F.: Pattern spectrum as a local shape factor for off-line signature verification. In: 13th ICPR, pp. 43–48 (1996)
12. Shekar, B.H., Bharathi, R.K.: Eigen-signature: A robust and an efficient off-line signature verification algorithm. In: ICRTIT, pp. 134–138 (2011)
13. Shekar, B.H., Bharathi, R.K., Sharmilakumari, M.: Kernel eigen-signature: An offline signature verification technique based on kernel principal component analysis. In: EACV 2011 Bilateral Russian-Indian Scientific Workshop, pp. 37–44 (2011)
14. Vargas, J., Ferrer, M., Travieso, C., Alonso, J.: Off-line signature verification based on grey level information using texture features. PR 44(2), 375–385 (2011)

A Robust On-road Moving Platform Video Stabilization Using Derivative Curve Warping

Deepika Shukla[1] and Rajib Kumar Jha[2]

[1] PDPM Indian Institute of Information Technology, Design and Manufacturing Jabalpur, India
[2] Indian Institute of Technology Patna, India

Abstract. In this paper a video stabilization technique based on derivative of integral projection curve has been proposed. The existing projection matching techniques consider only intensity values as the matching feature. Any change in pixel intensity values and their local distribution, results in shape variation of the projection curve. This shape variation affects the accuracy of curve matching and becomes a serious cause of wrong motion estimation. The proposed derivative curve technique considers the change in intensity values as the matching feature, instead of working on the intensity values itself. Inclusion of derivative provides shape information as high level feature and this results in better matching of two slightly dissimilar curves which are obvious for sequences taken from moving platform. Optimal matching of these derivative curves is achieved using Dynamic Time Warping technique.

Keywords: Integral projection, Dynamic time warping, Video stabilization.

1 Introduction

The videos captured using hand-held or moving platform camera are prone to undesired vibration induced motions. These undesired high frequency motions affect the visual quality of video sequence significantly. Video stabilization aims to remove the effect of such unwanted motions to provide a pleasant viewing experience. The relative motion between two consecutive frame is obtained using some motion estimation techniques. Accuracy of the estimated motion is highly desired as it directly reflects the performance of the whole stabilization system. In literature most of the motion estimation techniques are based on the block matching. The full block matching (BM) [1] technique provides accurate results in case pure translation but get affected by the presence of intensity variation or change in the local intensity distribution. It also suffers from large computational time. Robustness to intensity variation can be achieved using gray coded bit plane matching (GCBPM) instead of intensity frame and working on bit level reduces the processing time also [2]. Hybrid of these techniques have also been proposed to improve the accuracy of the motion estimation by working on a portion of frame to reduce the computational complexity [3]. Reducing the

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 343–348, 2013.

matching area always results in a compromise in terms of accuracy in comparison to the full frame techniques. The use of integral projection (IP) [4] reduces the matching problem from the 2D frame to 1D intensity projection thus inheriting the full frame information with reduced processing time requirement. Accuracy of estimated motion depend on the accurate matching of these curves. Matching criterion of all the discussed techniques is based on correlation to get some error minimization which gets affected by the change in local intensity distribution. An efficient algorithm for the projection curve matching has been proposed using dynamic time warping (DTW) providing optimal alignment of the two curves with better accuracy [5]. The DTW algorithm provides better alignment in case the two curves are similar in shape but a small acceleration or deacceleration and difference in intensity levels of the curve produces singularity in the optimal warping path. Such variations in the projection curves are very obvious in the videos captured from moving platform, which mostly suffer from jerks induced displacement along with the change in frame content caused by forward zoom-in effect.

In this paper the use of derivative projection-curve is proposed for robust alignment of two curves. The use of derivative provides inclusion of high level feature 'shape', instead of working directly on the intensity values. Comparison of the DTW based Curve Warping (CW) technique [5] is done with the derivative curve warping (DCW) in case of moving platform video. Performance evaluation of whole stabilization system is done in terms of computed PSNR values over entire image sequence.

2 Proposed Image Stabilization System

The DTW finds alignment by selecting the intensity value of the projection curve and finds its best match in the target curve. This intensity based matching may cause one point to be matched with multiple points having same intensity, thus producing pathological results in curve alignment. The DTW algorithm fails to find obvious natural alignment in case the two curves having peak or valley difference (i.e. intensity value) and shape change caused by small acceleration (i.e. expansion) or deacceleration (i.e. contraction). This alignment problem could be overcome by using a higher level feature 'shape' instead of using the intensity values as the matching feature. This shape information is obtained using first derivative of the projection curve. Derivative curve focuses on how the curve is changing instead of instantaneous curve point values [6].

2.1 Derivative Projection-Curve Extraction

The intensity projection curves for a frame of size $R \times C$ in the horizontal and vertical directions are obtained using (1).

$$\begin{aligned} P_H(j) &= \tfrac{1}{R} \textstyle\sum_{i=1}^{R} I(i,j); \forall j \in [1:C] \\ P_V(i) &= \tfrac{1}{C} \textstyle\sum_{j=1}^{C} I(i,j); \forall i \in [1:R] \end{aligned} \tag{1}$$

where $I(i,j)$ represents the pixel luminance value at location (i,j). Derivative curve (DC) of the estimated intensity projection can be obtained using left point estimation. DC along horizontal direction is given by (2).

$$DC_H(j) = P_H(i) - P_H(i-1); \quad where \quad 1 < i \leq C \quad and \quad j = i-1 \tag{2}$$

The obtained derivative curves of two consecutive frames are warped to find the optimal alignment of the respective intensity curves differing in their shape significantly.

2.2 Motion Estimation

Relative translational shift between two frames is clearly reflected in their respective projection curves. Adding the robustness in terms of variation in curve shape, instead of applying distance measure on the intensity curve, the difference of their derivative is proposed for better alignment. The distance matrix D representing the squared difference of derivative of the two horizontal curves DC_H of frame f and f-1 is given by (3).

$$D^f_{f-1}(i,j) = \quad (DC^f_H(i) - DC^{f-1}_H(j))^2 \tag{3}$$

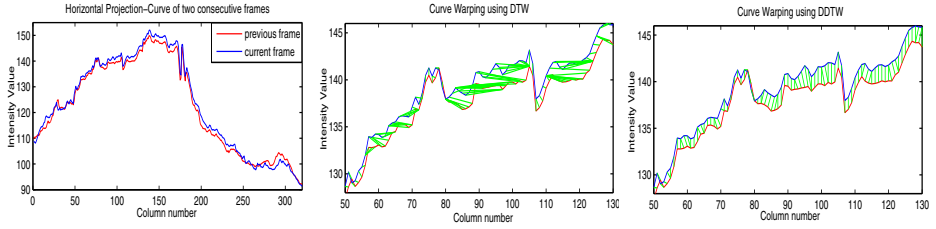
where $\forall i,j \in [1 : C-1]$ and in case of DC_V signature $\forall i,j \in [1 : R-1]$. To reduced the processing time and memory requirement, warping of the derivative curves is done using dynamic programming. in this algorithm an accumulated distance matrix AD is computed using recursive addition whose elements are given by (4).

$$\begin{aligned} AD(i,1) &= \textstyle\sum_{k=1}^{i} D(k,1); && \forall i \in [1 : C-1] \\ AD(1,j) &= \textstyle\sum_{k=1}^{j} D(1,k); && \forall j \in [1 : C-1] \\ AD(i,j) &= min[AD(i-1,j-1), AD(i-1,j), AD(i,j-1)] + D(i,j); \\ & && \forall i,j \in [2 : C-1] \end{aligned} \tag{4}$$

In AD matrix, the optimal warping path P_{opt} corresponds to best curve alignment and is extracted by backtracking the minimum valued element in its three adjacent neighbors. The optimal path P_{opt} starts with $P^L_{opt} = (C-1, C-1)$ and ends with $P^1_{opt} = (1,1)$. Other elements of optimal path P^l_{opt} can be obtained using (5):

$$P^l_{opt} := \begin{cases} (1, j-1); & if \quad i = 1, \\ (i-1, 1); & if \quad j = 1, \\ argmin\{AD(i-1, j-1), & \\ AD(i-1, j), AD(i, j-1)\}; & otherwise. \end{cases} \tag{5}$$

The P_{opt} obtained using proposed method provides best alignment and the uniqueness of path is ensured by following the lexicographical order in case the three neighbors have same value. To test the accuracy of proposed motion estimation algorithm, video 'CAR' is chosen. The sequence works as a representative



(a) (b) (c)

Fig. 1. (a) Shifted projection curves; Alignment using (b) CW (c) Proposed method

of real world moving platform videos. Figure 1(a) shows relative shift between two horizontal curves. To depict the effect of shape and intensity variation on accurate matching, the curve alignment using DTW based CW method and derivative DTW (DDTW) [6] based proposed DCW approach has been shown in Fig. 1(b) and Fig. 1(c) respectively. A portion of the two horizontal curves as shown in Fig. 1(a) has been selected for better visualization. Inclusion of shape information in the DDTW provides better alignment in comparison to the DTW, which suffers from the matching singularity (i.e. one to many point matching) at various location of curves.

2.3 Motion Correction and Compensation

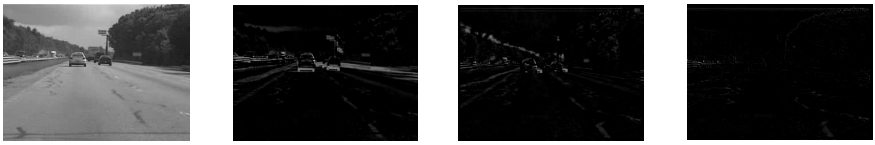
To get the motion estimation with respect to a reference frame, an absolute GMV ($AGMV$) for the each direction can be calculated using (6).

$$AGMV^f = AGMV^{f-1} + GMV^f \tag{6}$$

This $AGMV$ is used to separate the intentional panning from the jittery motion. Jitter being a high frequency oscillatory motion causes $AGMV$ to switch around certain value, while the intentional motion (panning) causes continuous shift in a single direction thus giving large motion vector values. Based on this concept a thresholding on the accumulated motion $AGMV$ can be used to separate out the two motions [1]. The reference frame is changed in case the AGMV crosses the preset motion threshold and future motion estimation is done w.r.t. new reference frame. The negative of these calculated $AGMV$s are used to compensate the effect of jitter by shifting the respective frame in opposite to the estimated direction.

3 Experimental Results

The moving platform videos suffer from the presence of large displacements caused due to platform jerks, and the forward motion of platform causes variation in the local pixel intensity distribution because of zoom-in effect. Existing



(a) (b) (c) (d)

Fig. 2. Stabilization measure on sequence *CAR* (a) 3^{rd} frame; Difference of 3^{rd} and 4^{th} frame (b) Original (c) CW (d) Proposed

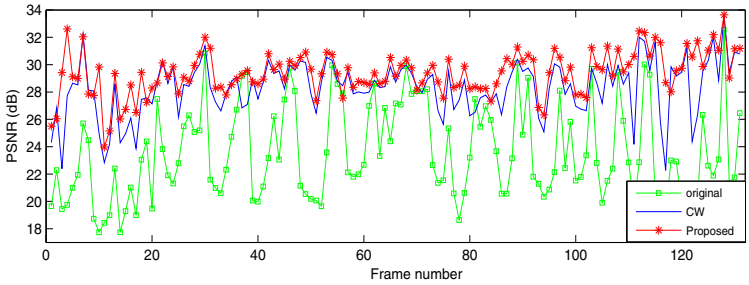


Fig. 3. Comparison of PSNR values for complete video *CAR*

intensity based stabilization techniques get affected by such variations and results in wrong motion estimation. The proposed DCW technique provides better stabilization due to accurate alignment of projection curves. In terms of motion accuracy, the CW [5] has earlier been suggested as the most efficient over the existing techniques. Efficiency of the proposed method over CW method for moving platform applications is evaluated using difference frame analysis, shown in Fig. 2. Fig. 2(a)-(d) shows the 3^{rd} frame of the video *CAR*, original difference frame of 3^{rd} and 4^{th} frame, corresponding stabilized results using CW and proposed DCW method respectively. The darker difference frame of DCW in comparison to CW method signifies better frame alignment. The proposed method is proved to be robust over CW in terms of accurate matching of projection curves. Comparison of the two methods is also done using estimated PSNR values corresponding to each pair of consecutive frame [3]. Proposed technique gives better PSNR as shown in Fig. 3. To validate the efficiency and robustness of the proposed method, various real-world moving platform videos, each of resolution 240×320 are chosen for performance analysis purpose. Fig. 4 shows the different test videos taken from moving bus. The performance evaluation is done in terms of interframe transformation fidelity (ITF) and processing time. ITF is an averaged PSNR value calculated between each consecutive pair of frames over entire video. Stabilized sequence results in a higher ITF value. Comparison of various stabilization techniques for different test videos is given in Table 1. Proposed technique proves to be less time efficient in comparison to GCBPM and IP methods but gives better stabilization over existing methods. The robustness of the proposed algorithm is achieved in terms of accurate motion estimation.

(a) (b) (c) (d)

Fig. 4. Other moving platform test video (a)Test1 (b)Test2 (c)Test3 (d) Test4

Table 1. Comparative results of the ITF and processing speed

Test Video	ITF						Processing time (sec/frame)				
	Orig.	Stabilized									
		BM	GCBPM	IP	CW	Proposed	BM	GCBPM	IP	CW	Proposed
Test1	21.71	21.94	22.65	22.89	23.16	24.33	2.124	0.107	0.059	40.72	0.148
Test2	20.10	21.35	23.32	24.10	24.73	26.89	2.190	0.102	0.070	40.11	0.137
Test3	20.44	20.83	21.69	23.57	24.03	24.74	2.018	0.092	0.069	38.96	0.128
Test4	22.48	23.71	24.01	25.63	25.97	27.64	2.316	0.130	0.063	40.61	0.156

4 Conclusion

This paper introduces a new derivative dynamic time warping based video stabilization technique. Robustness of the proposed technique has been discussed in the field of moving platform videos. The proposed DCW method provides better stabilization in terms of accurate projection curve alignment, providing higher ITF value in comparison to the existing intensity based stabilization techniques. Future work will focus on improving the time efficiency of the proposed algorithm to make it useful for real-time application.

References

1. Vella, F., Castorina, A., Mancuso, M., Messina, G.: Digital image stabilization by adaptive block motion vectors filtering. IEEE Transactions on Consumer Electronics 48(3), 796–801 (2002)
2. Ko, S.J., Lee, S.H., Jeon, S.W., Kang, E.S.: Fast digital image stabilizer based on gray-coded bit-plane matching. IEEE Transactions on Consumer Electronics 45(3), 598–603 (1999)
3. Li, S., Qi, J.: Image stabilization by combining gray-scale projection and representative point matching algorithms. In: 3rd International Conference on Awareness Science and Technology (iCAST 2011), pp. 128–132 (September 2011)
4. Piva, S., Zara, M., Gera, G., Regazzoni, C.S.: Color-based video stabilization for real-time on-board object detection on high-speed trains. In: IEEE Conference on Advantage Video and Signal Based Surveillance (AVSS 2003), pp. 299–304 (July 2003)
5. Bosco, A., Bruna, A., Battiato, S., Bella, G., Puglisi, G.: Digital video stabilization through curve warping techniques. IEEE Transactions on Consumer Electronics 54(2), 220–224 (2008)
6. Keogh, E.J., Pazzani, M.J.: Derivative dynamic time warping. In: The First SIAM International Conference on Data Mining (SDM 2001), pp. 1–11 (April 2001)

Efficient Image Plane Rotation Invariant Frequency Domain Face Recognition Technique Using Eye Localization

Papia Banerjee[1], Pradipta K. Banerjee[2], and Asit K. Datta[3]

[1] Department of Computer Science and Engineering
ABACUS Inst. of Engg. & Mgmt., Hoogly
papia.banerjee24@gmail.com
[2] Department of Electrical Engineering
Future Institute of Engineering and Management
Kolkata-700150
pradiptak.banerjee@gmail.com
[3] Department of Applied Optics and Photonics
University of Calcutta, Kolkata-700009
asitdatta@gmail.com

Abstract. An efficient frequency domain face recognition under arbitrary image plane rotations by single correlation filtering approach is proposed where faces are registered with eye detection. Here eye detection problem is carried out using shift invariant property of correlation filter. The proposed eye detection method includes log-polar transformation, correlation and regression neural network estimation. Proposed system shows the recognition improvement comparing to straightforward correlation filtering and multi-correlation approach.

1 Introduction

Frequency domain face recognition techniques are executed by cross-correlating the Fourier transform of test face image with a synthesized correlation filter(CF), generated from Fourier transform of training images and processing the resulting correlation output via inverse fast Fourier transform [4]. Though CFs are successfully used in face recognition,[4], the performance of CFs degrades drastically while the face images suffer from image plane rotation. In this case the standard CFs fail to detect the authentic person. To improve the performance of CFs it is then necessary to train the system with more images at certain intervals of angular positions, which causes degradation of signal-to-noise ratio. Further question may arise on what is the optimum number of rotated face images required for synthesizing one CF or how many synthesized CFs are enough for accurate recognition for arbritary rotation of face images. Designing a set of CFs for a single person implies multi correlation scheme in testing phase, which increases time complexity. To overcome these problems single correlation approach is carried out to recognize faces under any in-plane rotation for

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 349–354, 2013.

properly registered faces which is achieved by automatic localization of eyes. The eye detection process is carried out with log-polar transformation, correlation filter and neural network estimation. The effectiveness of the proposed method is verified with JAFEE database.

2 Proposed Face Registration Method

Fig.(1) shows the detail process of eye detection and face registration. The face image with any rotation ($0^o - 360^o$) is first log-polar transformed. A variant of unconstrained CF [2] termed as Eye-filter (EF) is designed with the eye templates, cropped from log-polar image in the training phase. The log-polar image is Fourier transformed and cross-correlated with EF. Eye is detected due to the shift invariant property of EF. The location of eye is found by searching the maximum value in the correlation plane. This location of one eye in log-polar is mapped into the original image for the both eyes positions by using the pixel coordinate information. The information of the pixel coordinates in two different domains is then fed to the generalized regression neural network (GRNN)[5]. In the testing stage this network automatically provides the actual position of eyes of rotated faces whenever the eye is detected in log-polar domain by the designed EF. With the actual eye coordinates the angular position of test face is determined and hence rotated in the opposite direction to register the face in its 0^o position.

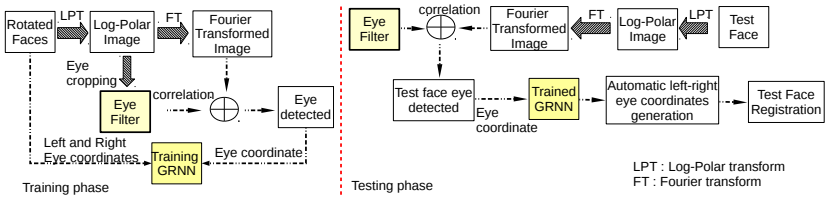


Fig. 1. Block diagram of detail process of face registration

3 Log-Polar Transformation and Eye Filter Synthesis

In addition to mapping in plane rotation into linear shift, log-polar transform offers some other features, such as wide field of view and a highly focused central area [3] as shown in Fig. 2. The measurement of linear shift of eyes in log-polar domain gives the information of the angular position of the face images in original scene. In case of unconstrained minimum average correlation energy (UMACE) [1] filter, conventionally, $\mathbf{x}_i$ (ith image) is used as an exemplar. However, instead of using $\mathbf{x}_i$, $(\mathbf{x}_i - \beta\mathbf{m})$, (where $\mathbf{m}$ represents mean image in Fourier domain) is introduced to modify the UMACE filter solution.

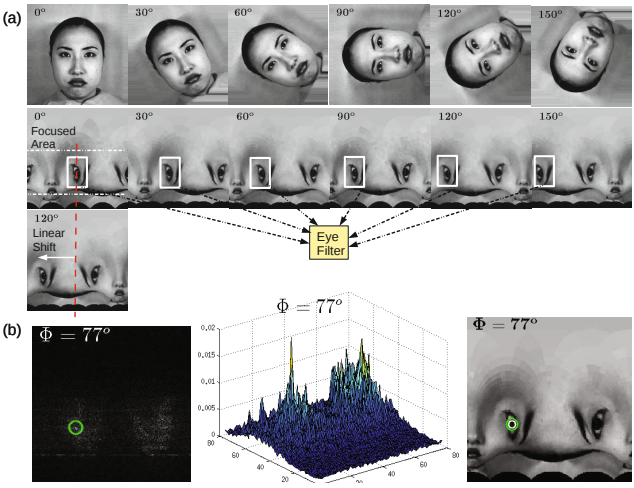


Fig. 2. (a) Log-polar transformation of rotated images and linear shift of eye. (b) Eye detection in log-polar image corresponding to maximum peak in correlation plane.

With this idea the average similarity measure (ASM) [1] can be expressed as

$$\text{ASM}_{new} = \frac{1}{N}\sum_{i=1}^{N}|\bar{\mathbf{X}}_i^*\mathbf{h} - (1-\beta)\bar{\mathbf{M}}^*\mathbf{h}|^2 = \mathbf{h}^+\bar{\mathbf{S}}_{new}\mathbf{h} \tag{1}$$

where

$$\bar{\mathbf{S}}_{new} = \frac{1}{N}\sum_{i=1}^{N}(\bar{\mathbf{X}}_i - (1-\beta)\bar{\mathbf{M}})(\bar{\mathbf{X}}_i - (1-\beta)\bar{\mathbf{M}})^* \tag{2}$$

Replacing $\bar{\mathbf{D}}$ of UMACE solution[1], by $\bar{\mathbf{S}}_{new}$ [2], the eye filter solution becomes

$$\mathbf{h}_{\text{EF}} = (\bar{\mathbf{S}}_{new})^{-a}(\bar{\mathbf{S}}_{new})^{-b}\mathbf{m} \tag{3}$$

Eq.(3) represents the eye filter (EF) synthesized with the eye templates as shown in Fig.(2). Having designed $\mathbf{h}_{\text{EF}}$ filter it is now correlated with the Fourier transformed log-polar image the shifted eye position will be detected and located by the maximum peak value on the resulting correlation surface. Fig.(2b) shows the correlation plane with a peak at desired location.

4 Estimating Eye Coordinates Using GRNN

Corresponding to the eye coordinate $P(u, v)$, shown in Fig.(3a) in log-polar image the two eye coordinates $E(x_1, y_1)$ and $E(x_2, y_2)$ of original image are manually collected in training stage. For different angle positions $(0^o, 30^o, 60^o, 90^o, 120^o, 150^o)$ of faces, a set of $P(u, v)$, $E(x_1, y_1)$ and $E(x_2, y_2)$ are obtained. These

values are fed to GRNN which can approximate any input-output map. Input layer of GRNN (Fig. 3b) is connected to pattern layer p_i which may contain q nodes, where q is the number of samples within a training data set. Each node represents the input vector, w_j , associated with the jth sample in training data. The signals of the pattern neurons i, are flowing into the third layer termed as summation layer. In this study $E(x_1, y_1)$ and $E(x_2, y_2)$ are used as target vectors where $P(u, v)$ is used for input vector w_i for training of GRNN. Given an unknown sample w', the approximation for Y can be obtained.

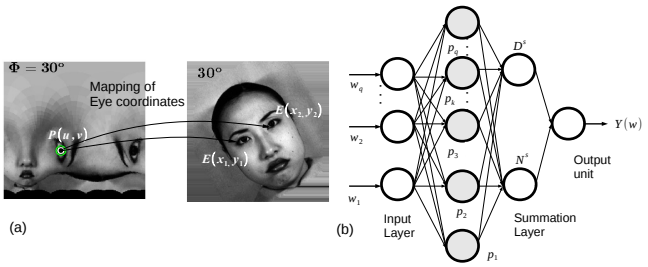


Fig. 3. (a) The mapping of eye coordinate in log-polar domain into original eye coordinates in Cartesian system is shown. (b) The typical GRNN model used for training.

5 Experimental Results

5.1 Eye Detection

Each image of JAFFE database has been manually rotated by angle in the range of $0 - 360^o$ with 1^o one degree interval, to get $360 \times 213 = 76,680$ number of images for testing purpose. This huge number of images are tested with the proposed method while the training data size contains only (a) 6 eye templates of one individual shown in Fig.(2) to synthesize the $\mathbf{h}_{\mathrm{EF}}$ filter and (b) 18 coordinates (3 for each image) set for training of GRNN. With the help of $\mathbf{h}_{\mathrm{EF}}$ and GRNN, eyes are found at any arbitrary angular position of test face image. Fig.(4) shows the eye localization of different images at different angular image plane rotations.

Having obtained the eye coordinates in Cartesian system the angular position of the test face is obtained as

$$\theta = \tan^{-1} \frac{y_2 - y_1}{x_2 - x_1} \tag{4}$$

where $\{x_i, y_i\}_{i=1,2}$ are the coordinates of eyes. After calculating θ the image is rotated by an angle $-\theta$ to get the retrieved face image.

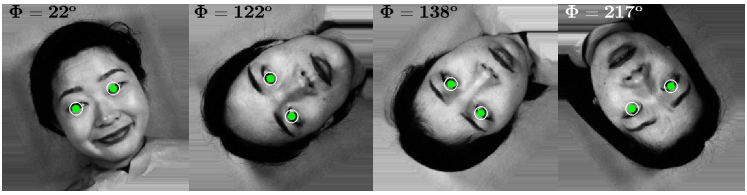


Fig. 4. Accurate eye localization of eyes at different angular positions of different images

5.2 Performance Analysis of Proposed Technique in Single Correlation Approach

At testing stage a standard CF[1] is used. This CF is synthesized with 5 images at 0^o position with image index $\{69, 72, 75, 78, 82\}$ of one person (shown in Fig.(4a)). One of the training images is now roated at an angle 70^o and then correlated with the designed CF. Fig.(5a) shows the poor PSR [4] value (5.13) obtained in this case. It is due to the fact the CF is designed at 0^o position and hence it could not recognize the image with 70^o rotation. Next the CF is again designed with the same images but at different angular positions. Each image is rotated from 0^o to 360^o with equal interval of 10^o. Hence one CF is designed

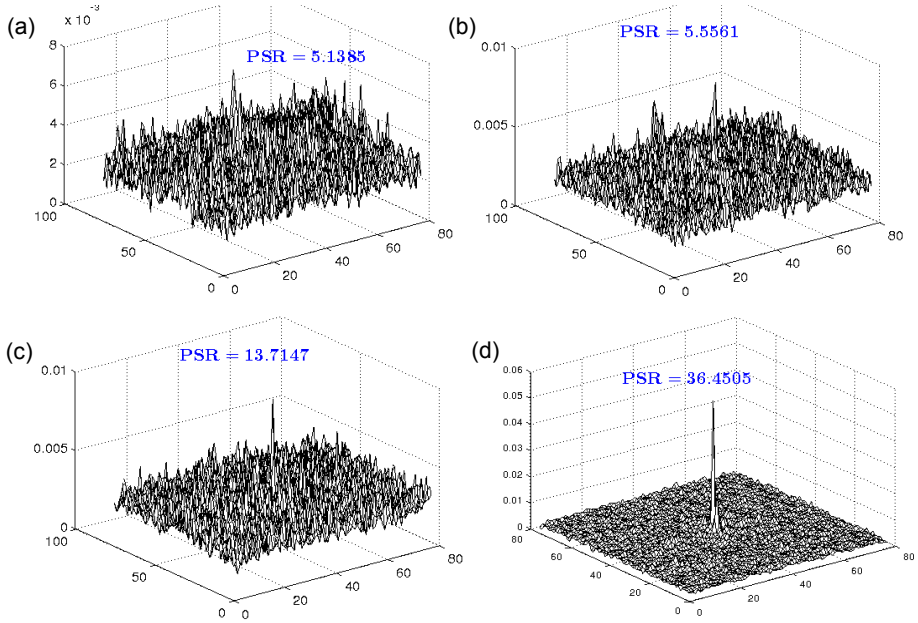


Fig. 5. Correlation planes in response to same test face for (a-b) different CF designs with angular positions (c) corresponds to multi-correlation approach (d) Proposed technique with single correlation approach

with 5×37 images. This CF is now tested with the image with 70^o rotation and the correlation output is shown in Fig.(5b).

Again poor PSR value (5.55) is obtained and no prominent peak is found. So increasing the size of the training set size does not resolve the problem here. Next the multi correlation approach is performed. Instead of one CF, for one person 4 CFs are designed. $0^o - 360^o$ rotated images are divided into 4 sub-angular divisions as $0 - 90^o$, $90 - 180^o$, $180 - 270^o$, $270 - 360^o$ and each subdivision contains 10^o intervals. With each angular subdvision one CF is synthesized. Such 4 CFs are correlated with the same test image and the corresponding correlation output is shown in Fig.(5c) corresponding to highest PSR value (13.71). Finally the same image with same angular condition is tested with the same CF (designed with only 5 mages at 0^o position) after face registration through eye detection. The resulting correlation output is shown in Fig.(5d) where a high PSR value (36.45) is obtained and sharp distinct peak observed at the correlation plane. In case of multi correlation approach 0.28 sec is elapsed and in case of the proposed technique 0.34 sec is consumed. But in term of PSR value and nature of correlation plane the proposed technique of single correlation approach outperforms the multi correlation approach.

6 Conclusions

This paper shows that the single correlation approach can be performed instead of multi correlation process for arbitrarily rotated (in-plane) faces by registering faces with eye detection. Proposed method considers only image plane rotations. How to extend the study in both illumination and pose variation along with image plane rotation needs further investigation.

References

1. Mahalanobis, A., Kumar, B.V., Song, S., Sims, S., Epperson, J.: Unconstrained correlation filter. Applied Optics 33, 3751–3759 (1994)
2. Banerjee, P.K., Datta, A.K.: Generalized regression neural network trained preprocessing of frequency domain correlation filter for improved face recognition and its optical implementation. Optics and Laser Technology 45, 217–227 (2013)
3. Bone, P., Young, R., Chatwin, C.: Position-, rotation-, scale-, and orientation-invariant multiple object recognition from cluttered scenes. Optical Engineering 45 (2006)
4. Kumar, B.V.K., Savvides, M., Xie, C., Venkataramani, K., Thornton, J., Mahalanobis, A.: Biometric verification with correlation filters. Applied Optics 43(2), 391–402 (2004)
5. Rutkowski, L.: Generalized regression neural networks in time-varying environment. IEEE Transaction on Neural Networks 15(3), 576–596 (2004)

A New Rotation Invariant Weber Local Descriptor for Recognition of Skin Diseases

Anabik Pal[1], Nibaran Das[2,*], Somenath Sarkar[3], Dwijendranath Gangopadhyay[4], and Mita Nasipuri[2]

[1] Department of Information Technology
[2] Department of Computer Science and Engineering
Jadavpur University, Kolkata (W.B), India
nibaran@gmail.com
[3] Department of Dermatology, School of Tropical Medicine, Kolkata (W.B), India
[4] Department of Dermatology, Burdwan Medical College, Burdwan (W.B), India

Abstract. A new rotation invariant Weber Local texture Descriptor(WLD) is proposed here. Performance of the developed features is evaluated on the basis of the recognition accuracies of SVM classifiers, trained and tested with the extracted features from the images of skins affected with three popular skin diseases such as Leprosy, Tineaversicolor, Vitiligo, and also normal skin, collected from School of Tropical Medicine, Kolkata. The WLD features are extracted with variations of the radius from 1 to 3 considering perimeters of having 8, 16 and 24 pixels respectively. The modified WLD provides an average improvement of 4.79% in recognition accuracy over the normal WLD in the present four class problem. Dividing each sample image into 4 sub-regions through its centre of gravity and extracting WLD features from each of them. Thus we have extracted two different feature sets having normal WLD and rotation invariant WLD. They provide the maximum recognition accuracies of 85.06% and 87.36% respectively using SVM classiifiers on test set.

Keywords: Texture, WLD, SVM, Leprosy, Tineaversicolor, Vitiligo.

1 Introduction

Nowadays different types of texture features are popularly used in computer vision problems in many areas such as industrial automation, biomedical image processing, remote sensing, biometric identification, face recognition, robot vision etc. In general, texture representations are mainly of three types namely Statistical, Spectral, and Structural. For statistical approaches different statistical moments are used to describe the texture of digital images. These approaches are based on Gray Level Co-occurrence matrix[1, 2], run length matrix[3], auto correlation, auto regression, MRF, moments etc. as found in literature. Spectral approach analyzes image texture in the frequency domain. As textures are periodic patterns, the entire frequency domain contains as much information as the image itself. Using this concept, several texture

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 355–360, 2013.

descriptors have been proposed. For structural approaches the basic scheme is to build a grammar for the texture and then parse the texture to see if it matches the grammar. The idea can be extended by defining texture primitives, simple patterns from which more complicated ones can be built. In that case, the parse tree for the pattern in a particular region can be used as a descriptor. Recently some local descriptors such as Local Binary Pattern(LBP)[4], Local Directional Pattern (LDP)[5], Weber Local Descriptor (WLD)[6] have been proposed. Among which, LBP is very popular compared to others. This is due to its easy implementation, short execution time, gray scale and rotation invariant property. Though WLD is more robust than LBP for noisy images, but the original WLD feature set is not rotation invariant. This motivated us to introduce rotation invariant property into the WLD feature set. Here the developed feature set is used to recognize three popular skin diseases such as Leprosy, Tineaversicolor and Vitiligo whose manifestations are all but similar in color. Later sample images are divided into four sub-regions by partitioning through it's center of gravity (CG) from where WLD features are extracted. This process emphasizes the local information in the feature set.

2 Weber Local Descriptor (WLD)

Weber's Law[7] states that the ratio of the increment threshold to the background intensity is a constant. This can be represented as

$$\frac{\Delta I}{I}=k \tag{1}$$

where ΔI is the increment threshold, which signifies the noticeable difference in 'I' for discrimination, 'I' represents the initial stimulus intensity, and 'k' is a constant. The fraction of the equation 1 is known as the Weber fraction. Using this law WLD has been proposed. It has two components 1) differential excitation (ξ) and 2) orientation (θ). For computing WLD for any pixel, it's neighborhood has to be defined first. The neighborhood is generally circular with a fixed number of pixels on the perimeter of the circle. Thus, the neighborhood is defined in terms of the radius R of the circle and the number of neighboring pixels P. Some typical neighborhoods are shown in Figure 1.

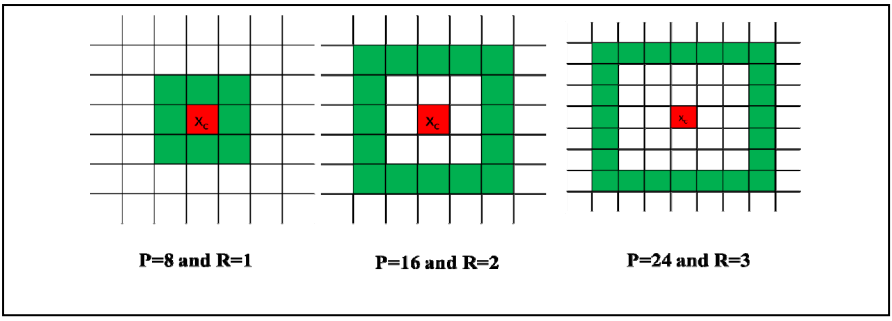


Fig. 1. Different neighborhood distribution

2.1 Differential Excitation (ξ)

This component tries to find the salient variations within an image to simulate the pattern perception of human beings. To compute differential excitation $\xi(x_c)$ of a current pixel x_c, the sum of differences of intensities between it and its neighbors is calculated by using following equation .

$$d_I^0 = \sum_{i=0}^{P-1} \Delta I = \sum_{i=0}^{P-1} \left[I(x_i) - I(x_c) \right] \tag{2}$$

where x_i (i=0, 1,……,P-1) denotes the i^{th} neighbors of x_c and P is the number of neighbors and $I(x_i)$ represents the intensity of the pixel x_i. Using Weber's Law, the ratio of the intensity differences to the intensity of the current pixel is then computed by using the following equation:

$$k(x_c) = \frac{d_I^0}{I(x_c)} \tag{3}$$

The differential excitation of the current pixel $\xi(x_c)$ is then computed as:

$$\xi(x_c) = \arctan\left[k(x_c)\right] = \arctan\left[\sum_{i=0}^{p-1} \frac{I(x_i) - I(x_c)}{I(x_c)}\right] \tag{4}$$

Obviously $\xi(x_c)$ takes negative value if the neighbors' intensities are smaller than that of the current pixel i.e. the surroundings are darker than the current pixel and takes positive values when the surroundings are lighter than the current pixel. To preserve this darker or brighter discriminating information $\xi(x_c)$ is not confined by taking absolute value of the numerator.

2.2 Orientation

The orientation component of WLD is computed as:

$$\theta(x_c) = \arctan\left(\frac{d_I^1}{d_I^2}\right) \tag{5}$$

where d_I^1 is the difference intensity between two diagonally opposite neighbors d_I^2 is the difference intensity between another pair of neighbors situated at the ends of a diagonal perpendicular to the previous one. For P=8 and R=1 the numerator and denominator are defined as below:

$$d_I^1 = I(x_5) - I(x_1) \tag{6}$$

$$d_I^2 = I(x_7) - I(x_3) \tag{7}$$

2.3 Rotation Invariant WLD Descriptor (WLDRI)

The differential excitation is rotation invariant as it computes the sum of intensity difference of center pixel with neighbor pixels so we get same value for any orientation of image pixels. But orientation component of WLD will not be fixed if the image is rotated. So, to make orientation component rotation invariant, all possible mutually perpendicular diagonal pairs are considered for computing θ_i as shown in equation 8 and out of them minimum one is chosen as shown in equation 9.

$$\theta_i = \arctan\left(\frac{I(x_{((P/2)+i) \bmod P}) - I(x_i)}{I(x_{((3P/4)+i) \bmod P}) - I(x_{((P/4)+i) \bmod P})}\right) \tag{8}$$

$$\theta(x_c) = \min_{i=0}^{P-1}(\theta_i) \tag{9}$$

Fig. 2. Diseases affected skin images

2.4 WLD Histogram

Obviously both Differential Excitation and Orientation can take values in between $-\pi/2$ and $+\pi/2$.The angle for Differential Excitation quantized into T equal width bins and for Orientation it is quantized into M equal width bins. Now we get a two dimensional histogram $h_{t,m}$. This histogram gives the feature of the image. Two pixel get same $h_{t,m}$ value when these two pixels differential excitations belong to the same bin, and orientations are also belonging to the same bin.

3 Experimental Setup

As mentioned earlier, we have used the developed rotation invariant WLD for the recognition of three popular skin diseases named as Leprosy, Tineaversicolor and Vitiligo which are all but similar in color while affected. The color changes due to formation of hypo-pigmented skin lesions in Leprosy, Tineaversicolor and formation of de-pigmented skin lesions in Vitiligo. The color of the lesions becomes lighter than the normal skin color for all three cases. We have collected the images of the affected regions from the outdoor patients of department of Dermatology, School of Tropical Medicine (STM) during the period April 2011- March 2012. During that period, we have managed to get images of the affected skin regions from 141 patients whose diseases were properly diagnosed by the dermatologists of the STM. The KODAK Easy share 143 camera having 3X optical Zoom and 12 megapixel resolutions has been used for capturing those images. The patients, already under treatment, was not considered in our study; only new cases were included. Multiple images were captured from the patients having more than one skin lesions. It is worthy to mention here that the images of normal skin of patients were also captured during the procedure. We have selected 876 images randomly from the collected images. The images are divided into two parts training and testing set maintaining nearly 4:1 ratio for each

of the four classes including normal skin. Some sample images of the affected regions are shown in Figure 2. The skin images are classified using a SVM based classification technique on the basis of WLD features (both WLD and WLDRI) extracted from those images. For WLD features, bins considers for $\xi(x_c)$ and $\theta(x_c)$ are 8 and 6 respectively. Here two variations in feature extraction process are considered as described below:

1) WLD features with fixed set of (P, R) values are extracted from the whole image.

2) Each image sample is partitioned into four regions by drawing a horizontal and a vertical line through the CG of the image. Here partitioning is done for giving emphasis on regional features extracted from the image samples and CG is chosen as the partitioning point for getting similar gray level distribution in all the four regions. Let the CG is the co-ordinate $C(c_x, c_y)$ then the equation for calculating CG of an image is given as below:

$$C(c_x, c_y) = \left(\frac{\sum_{i=0}^{M}\sum_{j=0}^{N} i.f(i,j)}{\sum_{i=0}^{M}\sum_{j=0}^{N} f(i,j)}, \frac{\sum_{i=0}^{M}\sum_{j=0}^{N} j.f(i,j)}{\sum_{i=0}^{M}\sum_{j=0}^{N} f(i,j)} \right) \tag{10}$$

Here the dimension of the image is M×N and f(i,j) is the intensity value at (i,j). WLD features with fixed set of (P,R) values are extracted from each of the four sub regions which are concatenated to form the feature set variations 1 and 2, mentioned above are repeated with different sets of values for P and R. Finally the features extracted in variation 3 are concatenated to form a new set.

4 Result and Discussion

The above mentioned features extraction techniques are applied to obtain different feature sets. The different sets of P and R values considered here are P=8 and R=1, P=16 and R=2, P=24 and R=3. The classification ability of these feature sets is tested by using multiclass SVM available in Weka 3.7.9 [8] tool. Train dataset is constructed here with 702 image samples and the rest 174 image samples are used as a test dataset for the present work. The recognition accuracy achieved with each feature extraction process is listed in Table1.We see that in all cases WLDRI is performed better than normal WLD and CG based partitioning increases the accuracy.

Table 1. Recognition accuracies obtained from different feature extraction techniques

No	Feature set used	Recognition Accuracy (in %)	
		Without Partition	CG Based Partition
1.	WLD T=8,M=6,P=8,R=1	72.98	74.14
2.	WLD T=8,M=6,P=16,R=2	74.71	78.16
3.	WLD T=8,M=6,P=24,R=3	71.84	85.06
4.	WLDRI T=8,M=6,P=8,R=1	74.71	78.74
5.	WLDRI T=8,M=6,P=16,R=2	79.31	82.76
6.	WLDRI T=8,M=6,P=24,R=3	79.88	87.36
7.	Concatenation of feature set 1, 2 and 3	83.91	86.78
8.	Concatenation of feature set 4, 5 and 6	85.06	**89.08**

5 Conclusion

In the present work, a new rotation invariant Weber Local texture descriptor (WLD) is proposed to identify three similar looking skin diseases such as Leprosy, Tineaversicolor and Vitiligo on the basis of the performance of SVM classifier. The developed feature performs better than the original WLD and provides an average improvement of 4.79% in recognition accuracy in the present four class problem. We have also proposed a CG based region partitioning scheme to divide sample images into 4 subregions for extraction of more local information from every sub-regions. This provides the maximum recognition accuracy of 85.06% and 87.36% for normalfor normal WLD and rotation invariant WLD respectively. Multiscale analysis of both the features are also done here by choosing three sets of (P, R) values and a highest recognition accuracy of 89.08% is achieved with WLDRI and CG based partitioning. The obtained recognition accuracy is quite satisfactory. Introduction of stronger feature set and/or with different classifier combination or fusion techniques may be explored further.

Acknowledgment. Authors are grateful to the "Center for Microprocessor Application for Training Education and Research" of Computer Science & Engineering Department, Jadavpur University laboratory, for providing infrastructure facilities during progress of the work. Authors are also thankful to Department of dermatology, School of Tropical medicine, Kolkata for providing patients and medical expertise during data collection drive.

References

[1] Haralick, R.M., et al.: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics SMC-3, 610–621 (1973)
[2] Haralick, R.M.: Statistical and structural approaches to texture. Proceedings of the IEEE 67, 786–804 (1979)
[3] Galloway, M.M.: Texture analysis using gray level run lengths. Computer Graphics and Image Processing 4, 172–179 (1975)
[4] Ojala, T., et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 971–987 (2002)
[5] Jabid, T., et al.: Gender Classification using Local Directional Pattern (LDP). Presented at the International Conference on Pattern Recognition (2010)
[6] Jie, C., et al.: WLD: A Robust Local Image Descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 1705–1720 (2010)
[7] Jain, A.K.: Fundamentals of Digital Signal Processing: Prentice-Hall (1989)
[8] Hall, M., et al.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11, 10–18 (2009)

A Composite Wavelets and Morphology Approach for ECG Noise Filtering

Vikrant Bhateja[1], Shabana Urooj[2], Rini Mehrotra[1], Rishendra Verma[1], Aimé Lay-Ekuakille[3], and Vijay Deepak Verma[4]

[1] Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow (U.P.), India
[2] Department of Electrical Engineering, School of Engineering, Gautam Buddha University, Greater-Noida (U.P.), India
[3] Department of Innovation Engineering, University of Salento, Lecce, Italy
[4] Medical University & Associated Hospitals Greater Noida UP India
{bhateja.vikrant,rini684.mehrotra,verma.rishendra90, shabanabilal}@gmail.com

Abstract. Noisy ECG signals contain variations in the amplitudes or in the time intervals which represents the abnormalities associated with the heart; thereby making visual diagnosis difficult for cardiovascular diseases. Hence, to facilitate proper analysis of ECG; this paper presents a combination of wavelets analysis and morphological filtering as an approach for noise removal in ECG signals. The proposed algorithm involves sub-band decomposition of ECG signal using bi-orthogonal wavelet family. The wavelet detail coefficients containing the noisy components are then processed by morphological operators using linear structuring elements. The morphological filter processes only the corrupted bands without affecting the signal parameters. Simulation results of the proposed algorithm show noteworthy suppression of noise in terms of higher signal-to-noise ratio preserving the ST segment and R wave of ECG.

Index Terms: bi-orthogonal wavelets, detail coefficients, morphological filtering.

1 Introduction

Electrocardiogram (ECG) is measure of the voltage fluctuations occurring inside heart because of the potential changes across the cell membrane. During acquisition from the patient's body, ECG signal gets distorted by different types of noise artifacts that can be within a frequency band of interest. ECG signals are plagued by impulse noise, due to muscle activities; it appears as false positive and negative voltage fluctuations (in the form of spikes of short duration) in the ECG signal [1-2]. Therefore, ECG signal conditioning for providing noise removal is a necessary requirement for computer-aided analysis such as QRS detection and temporal alignment. Computer based diagnosis have been demonstrated fertile for various other diseases like breast cancers

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 361–366, 2013.

[3]-[6], brain tumors [7]-[8] as well as for diseases based on accumulation of fluid [9]-[12]. In this regard, the reliability issues of the computer-based analysis models reported in former works are of great importance. Numerous variants of morphological filtering operations have been proposed in literature for noise filtering of ECG signals. Processing of ECG signals with different variants of morphological operators employed flat structuring elements were able to reduce ST segment distortion only up to 50% [13]-[15]. Wavelet based techniques [16]-[18] for ECG noise suppression requires numerous experiments for ruling out scales and thresholds. These algorithms were complex incorporating mother wavelet selection and may require more computation time. R. Verma *et al*. [19]-[21] proposed another method for impulse noise suppression and baseline correction in ECG employing three as well as two dimensional structuring elements. In the present work, an improved filtering technique using a combination of wavelets and morphology is employed for achieving low distortion and high noise suppression. The rest of this paper is organized as follows. The proposed ECG filtering approach is described in Section 2. Section 3 provides the obtained simulation results and discusses them while the work is concluded under Section 4.

2 Proposed ECG Filtering Approach

ECG signal being a non-stationary signal makes it difficult for denoising. A proficient technique for such a non-stationary signal processing is the wavelet transform. The wavelet transform can be used for decomposition of a signal in the time frequency scale plane. Morphological filtering is a non-linear transformation technique primarily used for local modification of geometrical features of a signal. This shape information (of features) is extracted by using a structuring element (of appropriate dimensions) to operate on the input signal. Erosion, dilation, opening and closing are the common morphological operators used. [22]-[23]. An important aspect in noise suppression is the selection of optimum size of the structuring element. Improper selection of structuring element may distort the adjacent wave in the ECG signal. Moreover, the ECG noise filtering by pure morphological filter leaves residual impulse noise which brings variations in the amplitude and time interval of the processed ECG signal. However, biorthogonal wavelet provides frequency domain approach for noise filtering. It is noteworthy that decomposition of ECG signal into detail sub-bands helps in dealing with the residual noise which cannot be removed through morphological filter. Hence, in this work, combination of wavelet and morphology is used for noise filtering of ECG signal. The signal is first decomposed into subbands; the wavelet detail coefficients are then processed by applying morphological operations using linear structuring element. The proposed algorithm comprises of following steps:

Step-1: The algorithm initiates with the decomposition of the ECG signal into subbands. This is done by choice of appropriate wavelet family like Bi-orthogonal wavelets for ECG signal processing. Most of the impulse noise which is characterized by

high spikes resides in the detail sub-bands. Thus, the detail sub-bands with high frequency spikes are separated for application of proposed morphological filter.

Step-2: The morphological filter involves the application of combination of erosion and dilation operators on the affected sub-bands. ECG signal (A) is mainly corrupted by impulses (or spikes) i.e. very large positive or negative values of very short interval. The noise suppression with the proposed morphological filter involves dilation followed by erosion in the first stage and erosion followed by dilation in second stage using linear structuring element (B). The structuring element should be comparable to the width of impulse noise. The linear structuring element B used in this work consists of an array of ones of size 10. This enables removal of residual impulse noise without introducing any distortion in ECG waveform or loss of information. The results of both the stages are averaged as given by equation (1) to get the denoised detail sub-bands.

$$B = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1] \tag{1}$$

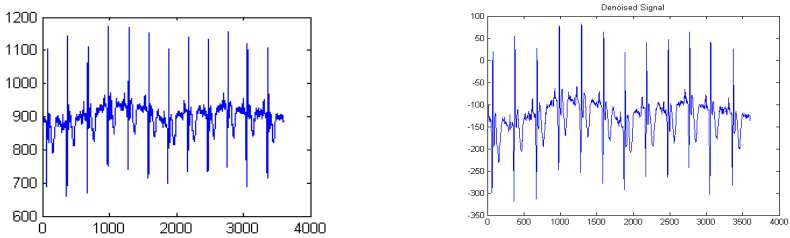
$$f = \frac{((A\theta B) \oplus B) + ((A \oplus B)\theta B)}{2} \tag{2}$$

Where: B is a structuring element, f represents the reconstructed ECG signal, θ represents erosion operation, $\oplus$ represents dilation operation.

Step-3: The denoised output ECG signal is finally reconstructed from the the processed detail sub-bands through wavelet reconstruction.

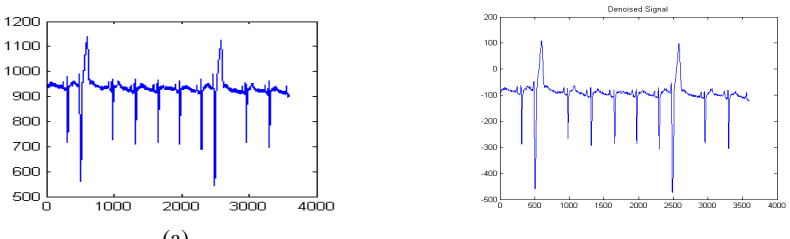
3 Results and Discussions

For simulation purposes in the present work, ECG signals are adapted from internationally accepted MIT-BIH database that consists of recordings which are observed in clinical practice. The ECG recordings were created using two clean recordings (118 and 119) from the MIT-BIH Arrhythmia Database, to which calibrated amounts of noise from record 'em' were added which are termed as Noise Stress Databases [24]. ECG signal interpretation is based on the wave amplitude, time interval and the shape of QRS complexes. These signal parameters are affected by the noise which interfere with the actual information and thereby makes visual diagnosis difficult [25]. It can be noticed that in the fig. 1(a) and 2(a) the QRS complex is corrupted from noise artifacts. The corrupted ECG signal is first decomposed into sub-bands by choice of appropriate wavelets. The noise content in the detail sub-bands is then suppressed by using morphological filtering. The choice of optimal structuring element is experimentally determined by using the Signal to Noise Ratio (SNR) as performance measure for quality evaluation [26]-[27]. It can be seen from fig. 1(b) and 2(b) that impulse noise is reduced by the proposed approach. Evaluation of SNR values will act as tool for the comparison of different families. The SNR value of noisy signals 118e24, 119e24 and PTB s0010 were found to be 25.21 dB, 36.27 dB and 24.74 dB respectively. Upon denoising with the proposed approach, SNR values get enhanced to 32.2 dB, 44.5dB and 26.5 dB respectively.



(a) (b)

Fig. 1. (a) Noisy ECG Signal (MIT BIH record 118e24), (b) Denoised ECG Signal using proposed approach



(a) (b)

Fig. 2. (a) Noisy ECG Signal (MIT BIH record 119e24), (b) Denoised ECG Signal using proposed approach

It has been experimentally analyzed that the biorthogonal wavelets appears most promising among all as it gives high quality results on all the signals evaluated by the proposed algorithm. Higher value of SNR obtained by biorthogonal wavelets depicts better noise suppression in comparison to other wavelet families. The application of morphology with wavelets overcomes the need of wavelet thresholding.

4 Conclusion

This paper presents the decomposition of the ECG signal using wavelets followed by noise filtering through morphology. The ECG signal decomposition helps in multi-scale analysis which provides noise filtering only in the corrupted sub-bands. Noise suppression is achieved through morphological filtering using linear structuring elements. This enables proper removal of residual impulse noise by applying morphological filter only in the affected sub-bands without introducing any distortion in ECG waveform. The proposed algorithm provides good results by bi-orthogonal wavelet in comparison to other wavelet families in terms of SNR values.

References

1. Dupre, A., Vincent, S., Iaizzo, P.A.: Basic ECG Theory, Recordings and Interpretation. In: Handbook of Cardiac Anatomy, Physiology, and Devices, pp. 191–201 (2005)

2. Gupta, R., Bera, J.N., Mitra, M.: Development of An Embedded System and MATLAB-Based GUI for Online Acquisition and Analysis of ECG Signal. Journal Measurement 43(9), 1119–1126 (2010)
3. Jain, A., Singh, S., Bhateja, V.: A Robust Approach for Denoising and Enhancement of Mammographic Breast Masses. International Journal on Convergence Computing 1(1), 38–49 (2013)
4. Bhateja, V., Urooj, S., Pandey, A., Misra, M., Lay-Ekuakille, A.: A Polynomial Filtering Model for Enhancement of Mammogram Lesions. In: Proc. of IEEE International Symposium on Medical Measurements and Applications (MeMeA 2013), Gatineau, Quebec, Canada, pp. 97–100 (2013)
5. Bhateja, V., Urooj, S., Misra, M., Lay-Ekuakille, A.: A Robust Polynomial Filtering Framework for Mammographic Image Enhancement from Biomedical Sensors. IEEE Sensors Journal 13(11), 4147–4156 (2013)
6. Bhateja, V., Urooj, S., Pandey, A., Misra, M., Lay-Ekuakille, A.: Improvement of Masses Detection in Digital Mammograms employing Non-Linear Filtering. In: Proc. of IEEE International Multi-Conference on Automation, Computing, Control, Communication and Compressed Sensing (iMac4s 2013), pp. 406–408 (March 2013)
7. Srivastava, A., Alankrita, R.A., Bhateja, V.: Combination of Wavelet Transform and Morphological Filtering for Enhancement of Magnetic Resonance Images. In: Snasel, V., Platos, J., El-Qawasmeh, E. (eds.) ICDIPC 2011, Part I. Communications in Computer and Information Science, vol. 188, pp. 460–474. Springer, Heidelberg (2011)
8. Raj, A., Alankrita, A.S., Bhateja, V.: Computer Aided Detection of Brain Tumor in Magnetic Resonance Images. International Journal of Engineering and Technology (IACSIT) 3(5), 523–532 (2011)
9. Urooj, S., Lay-Ekuakille, A., Ansari, A.Q., Khan, M., Vergallo, P., Trotta, A.: Volumetric estimation of thorax with cylindrical model and anthropometric measurements. In: 2011 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA), pp. 213–216. IEEE (May 2011)
10. Urooj, S., Khan, M., Ansari, A.Q.: Thorax Physiological Monitoring and Modeling for Diagnosis of Pulmonary Edema. International Journal of Measurement Technologies and Instrumentation Engineering (IJMTIE) 1(2), 54–60 (2011)
11. Urooj, S., Khan, M., Ansari, A.Q., Salhan, A.K.: Measurement of Thoracic Impedance and Approximations: A Diagnosis Technique for Clinical Utilization. Indian Journal of Industrial and Applied Mathematics 3(2), 85–93 (2012)
12. Urooj, S., Khan, M., Ansari, A.Q., Lay-Ekuakille, A., Salhan, A.K.: Prediction of quantitative intrathoracic fluid volume to diagnose pulmonary oedema using LabVIEW. Computer Methods in Biomechanics and Biomedical Engineering 15(8), 859–864 (2012)
13. Sun, Y., Chan, K.L., Krishnan, S.M.: ECG Signal Conditioning by Morphological Filtering. Computers in Biology and Medicine 32(6), 465–479 (2002)
14. Chu, C.-H.H., Delp, E.J.: Impulsive Noise Suppression and Background Normalization of Electrocardiogram Signal Using Morphological Operators. IEEE Transactions on Biomedical Engineering 36, 262–273 (1996)
15. Liu, Z., Wang, J., Liu, B.: ECG Signal Denoising Based on Morphological Filtering. In: Proc. of the 5th International Conference on Bioinformatics and Biomedical Engineering, Wuhan, pp. 1–4 (May 2011)
16. Sayadi, O., Shamsollahi, M.: Multiadaptive Bionic Wavelet Transform: Application to ECG De-noising and Baseline Wandering Reduction. EURASIP Journal on Advances in Signal Processing 2007, 1–11 (2007)

17. Alfaouri, M., Daqrouq, K.: ECG Signal De-noising by Wavelet Transform Thresholding. American Journal of Applied Sciences 5(3), 276–281 (2008)
18. Sharma, L.N., Dandapat, S., Mahanta, A.: ECG Signal denoising Using Higher Order Statistics in Wavelet Subbands. Biomedical Signal Processing and Control 5, 214–222 (2010)
19. Verma, R., Mehrotra, R., Bhateja, V.: An Improved Algorithm for Noise Suppression and Baseline Correction of ECG Signals. In: Satapathy, S.C., Udgata, S.K., Biswal, B.N. (eds.) Proc. of Int. Conf. on Front. of Intell. Comput. AISC, vol. 199, pp. 733–739. Springer, Heidelberg (2013)
20. Verma, R., Mehrotra, R., Bhateja, V.: A New Morphological Filtering Algorithm for Pre-Processing of Electrocardiographic Signals. In: Proc. (SPRINGER) of the Fourth International Conference on Signal and Image Processing (ICSIP 2012), Coimbatore, India, pp. 193–201 (December 2012)
21. Mehrotra, R., Verma, R., Bhateja, V.: An Integration of Improved Median and Morphological Filtering Techniques for ECG Processing. In: Proc. of the 3rd International Advance Computing Conference (IACC 2013), Ghaziabad, U.P., India, pp. 1212–1217 (2013)
22. Bhateja, V., Devi, S.: A Novel Framework for Edge Detection of Microcalcifications using a Non-Linear Enhancement Operator and Morphological Filter. In: Proc. of IEEE 3rd International Conference on Electronics & Computer Technology (ICECT 2011), Kanyakumari, India, vol. 5, pp. 419–424 (April 2011)
23. Lay-Ekuakille, A., Vergallo, P., Griffo, G., Conversano, F., Casciaro, S., Urooj, S., Bhateja, V., Trabacca, A.: Mutidimensional Analysis of EEG Features using Advanced Spectral Estimates for Diagnosis Accuracy. In: Proc. of IEEE International Symposium on Medical Measurements and Applications (MeMeA 2013), Gatineau, Quebec, Canada, pp. 237–240 (May 2013)
24. Moody, G.B., Mark, R.G.: Impact of the MIT-BIH Arrhythmia Database. IEEE Engineering in Medical and Biology Magazine 20, 45–50 (2001)
25. Jain, A., Bhateja, V.: A Novel Image Denoising Algorithm for Suppressing Mixture of Speckle and Impulse Noise in Spatial Domain. In: Proc. of IEEE 3rd International Conference on Electronics and Computer Technology (ICECT), Kanyakumari, India, vol. 3, pp. 207–211 (2013)
26. Jain, A., Bhateja, V.: A Full-Reference Image Quality Metric for Objective Evaluation in Spatial Domain. In: Proc. of IEEE International Conference on Communication and Industrial Application (ICCIA), Kolkata, W. B., India, pp. 91–95 (2011)
27. Singh, S., Jain, A., Bhateja, V.: A Comparative Evaluation of Various Despeckling Algorithms for Medical Images. In: Proc. of (ACMICPS) CUBE International Information Technology Conference & Exhibition, Pune, India, pp. 32–37 (2012)

A Bi-level IHS Transform for Fusing Panchromatic and Multispectral Images

Navaneeth K. Ramakrishnan and Philomina Simon

Department of Computer Science, University of Kerala, Thiruvananthapuram
{navaneeth.kr,philomina.simon}@gmail.com

Abstract. In this paper, a new method for Panchromatic and Multispectral satellite image fusion is proposed. The major challenge of a fusion algorithm is to improve the spatial and spectral qualities of the fused image. But the spatial and spectral qualities are inversely proportional; we cannot improve either quality above particular range without losing visual quality, and most of the current methods do not take into consideration about visual quality. The proposed method tries to improve the spatial and visual quality with reduced spectral distortion using a Bi-Level Intensity Hue Saturation transform. Proposed method is rigorously tested over QuickBird and IKONOS satellite images and the experimental results shows that our method produces high visual quality fused images with a good spatial and spectral quality levels compared with existing methods.

Keywords: Bi-Level IHS, Image Fusion, QuickBird, IKONOS, Visual Quality, Remote Sensing.

1 Introduction

Satellite images are widely used in remote sensing applications. Earth observing satellites like QuickBird and IKONOS provide two types of images; Multispectral (MS) and Panchromatic (PAN) images. MS images have high spectral resolution and PAN images have very good spatial resolution. It is desirable to have the best possible spatial resolution in order to find fine features on the Earth's surface for most remote sensing applications. Moreover, a high spectral resolution is also required to discriminate among different ground covers. Most of these remote sensing applications need human intervention for feature detection and classification of land area, so that visual quality of the fused image is also an important factor. However, for certain applications, there is a need to improve the spatial resolution of the MS images. Improving the spatial quality of the MS image is usually called pansharpening. The main issue in the pansharpening is the inverse proportionality between spectral and spatial quality. When we try to improve one quality the other quality become reduced. A good fusion algorithm must attain a good spectral and spatial quality level, that is, good spatial quality with less spectral distortion. Mapping and surveillance application demand not only high spatial and spectral quality images but also the good visual quality for images. Most of the methods are not concerned about the visual quality of the fused image.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 367–372, 2013.

Within the well known spatial domain fusion methods IHS[1], Brovey[2] and PCA[2] methods achieve fine spatial quality but generate spectral distortions. Another family of methods using wavelets produces much less spatial quality images than those fused by the Brovey, IHS, and PCA methods. Some of the new transformations such as courvelet, contourlet[3][4] and NSCT[5][6] are proposed to overcome this weakness. The contourlet has many advantages over curvelet. A. G. Mahyari *et al*[5]., proposed a new method for identifying a best threshold for fusing MS and PAN images. But algorithm failed to provide good spatial quality. Then researchers try to combine spatial and transform domain fusion methods. Hybrid methods like NSCT and IHS[6], Bilateral filter and IHS[7] are try to combine good qualities of both techniques, and able to achieve good spectral quality but the sharpness of the fused image is lost, also the time complexity is also very high.

Many new less complex methods are introduced to reduce the spectral distortion happened during the image fusion using IHS transform. Within that Adaptive-IHS[8] and Improved-IHS[9] provide good spectral quality but the image is visually distorted. The new IHS fusion by Choi[10] able to preserve the spectral and spatial qualities up to a good level, but lose of sharpness reduces the visual quality. In this paper out objective is to achieve a good visual quality by fusing PAN spatial details along with color details of MS.

2 Bi-level IHS Fusion Algorithm

A good fusion algorithm should improve two inversely proportional qualities, ie., spatial and spectral, which is not possible simultaneously. So we are trying to achieve a good spatial and spectral quality levels along with good visual quality using a Bi-Level IHS algorithm. The Bi-Level IHS method has two levels of IHS based fusion algorithms, within that the first level algorithm brings spatial information to the fused image using fast IHS and new adaptive averaging; and the second level algorithm fuse spectral details to the MS image.

The first level of proposed fusion method uses a histogram equalized PAN image for the fast IHS transform, and that can be implemented as follows,

$$\mathrm{PAN_{eq}} = \mathrm{PAN} * \frac{\sigma_{\mathrm{I}}}{\sigma_{\mathrm{PAN}}} \tag{1}$$

where, σ_{I} standard deviation of intensity component and σ_{PAN} is the standard deviation of PAN image.

$$\begin{bmatrix} \mathrm{R'_{IHS}} \\ \mathrm{G'_{IHS}} \\ \mathrm{B'_{IHS}} \end{bmatrix} = \begin{bmatrix} \mathrm{R} + (\mathrm{PAN_{eq}} - \mathrm{I}) \\ \mathrm{G} + (\mathrm{PAN_{eq}} - \mathrm{I}) \\ \mathrm{B} + (\mathrm{PAN_{eq}} - \mathrm{I}) \end{bmatrix} \tag{2}$$

$$\mathrm{I} = (\mathrm{R} + \mathrm{G} + \mathrm{B} + \mathrm{NIR})/4 \tag{3}$$

The next step in first level operation is to apply the adaptive averaging function. The adaptively added MS components and PAN details helps to improve the spatial quality and reduces the spectral distortion. The averaging function is:

$$\begin{bmatrix} R''_{IHS} \\ G''_{IHS} \\ B''_{IHS} \end{bmatrix} = \begin{bmatrix} (R'_{IHS} + PAN_{eq} + R * \alpha_1)/(\alpha_1 + 2) \\ (G'_{IHS} + PAN_{eq} + G * \alpha_2)/(\alpha_2 + 2) \\ (B'_{IHS} + PAN_{eq} + B * \alpha_3)/(\alpha_3 + 2) \end{bmatrix} \quad (4)$$

$$\alpha_1 = \mathrm{corr}(R, PAN) \quad (5)$$

$$\alpha_2 = \mathrm{corr}(G, PAN) \quad (6)$$

$$\alpha_3 = \mathrm{corr}(B, PAN) \quad (7)$$

The Second Level IHS transform add color details to the fused images. Color information is an important factor while we considering the visual quality of the image. Also to improve the spectral quality the original Hue and Saturation components of the MS image is fused with the intensity component of the previously fused bands$\{R''_{IHS}, G''_{IHS}, B''_{IHS}\}$. The fused bands transformed HIS[11] domain,

$$\begin{bmatrix} I'' \\ H'' \\ S'' \end{bmatrix} = \mathrm{HSI}(R''_{IHS}, G''_{IHS}, B''_{IHS}) \quad (8)$$

Use Hue and Saturation of the original MS images and the intensity component of the fused image are used to find the final fused MS images.

$$\begin{bmatrix} R_f \\ G_f \\ B_f \end{bmatrix} = \mathrm{inverse\_HSI}(I'', H, S) \quad (9)$$

3 Experimental Analysis

To evaluate our method we selected QuickBird and IKONOS satellite images which have 4m resolution for MS bands and 1m resolution for PAN image. For testing the quantization level reduced from 11 bits to 8 bits. Since there are no reference images are available, we use the original MS images as reference images and scale down the fused MS images and PAN image to the size of reference image for the quantitative analysis.

In order to evaluate the performance and effectiveness of our proposed fusion method, we compared the new algorithm with five other existing fusion methods; IHS, modified IHS by Choi, NSCT+ IHS, Improved-IHS and Adaptive IHS. From the images in Fig. 1 and Fig. 2 we can observe that the IHS produces the color distortions. The modified IHS, IKONOS IHS and improved IHS by Choi are fused the images with less spectral distortions but the spatial and visual qualities are very poor. The NSCT and IHS combined method produce low spatial and visual quality images but the spectral quality is good, also the computation time for the fusion process is very high as compared with all other methods. Proposed method produce good spatial and visual quality images with less spectral distortions than IHS. To provide quantitative comparison of different fusion algorithms, we use eight objective indicators shown in Table 1.

Here SSIM[12] is used as a quantitative measure for the visual quality. The PAN image is considered as the perfect quality image. The RGB image is converted to grayscale and compared with the PAN image. When more structural details are preserved in the fused image; better the visual quality. Spectral distortion is calculated using three different measures; ERGAS[13], RASE[14] and SAM, while spatial quality is measured using Cross Correlation (CC) and Universal Image Quality Index (UIQI)[15]. Cross

correlation and UIQI are calculated for each band with PAN image. In order to analyze the performance of our method is achieve a good spectral and spatial quality levels along with visual quality, the sum of difference between the spatial quality improvement and spectral distortions, and SSIM is calculated, shown in Table 1, Eq. No 8. Since RASE is calculated as rate of distortion; CC and SSIM is calculated as the rate of spatial and visual quality improvement, higher values of the QDIFF will represent the good performance of the algorithm. Table 2, shows the quantitative analysis of the IKONOS and QuickBird images.

Table 1. Quantitative metrics

Eq. No	Quantity	Equation	Parameters
1	RASE	$RASE = \frac{100}{M}\sqrt{\frac{1}{N}\sum_{i=1}^{N} RMSE^2(B_i)}$	M is the mean radiance of the N spectral bands, (Bi) of the original MS bands. , $RMSE^2(B_i) = bias^2(B_i) + SD^2(B_i)$
2	ERGAS	$ERGAS = 100\frac{h}{l}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\frac{RMSE^2(B_i)}{M_i^2}}$	h is the resolution of PAN, l is the resolution of MS, M_i is the mean radiance of each spectral band
3	SAM	$SAM(v,\hat{v}) = arccos\left(\frac{\langle v,\hat{v}\rangle}{\|v\|_2\,\|\hat{v}\|_2}\right)$	Where $v = \{v_1, v_2, \ldots, v_L\}$is the original spectral pixel vector $\hat{v} = \{\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_L\}$is the distorted pixel vector
5	CC	$\sum_{i=1}^{M}\sum_{i=1}^{N}\frac{(A(i,j)-\mu_A)(B(i,j)-\mu_B)}{\sqrt{(\sum_{i=1}^{M}\sum_{i=1}^{N}(A(i,j)-\mu_A))^2(\sum_{i=1}^{M}\sum_{i=1}^{N}(B(i,j)-\mu_B))^2}}$	where μ_A and μ_B are the mean, σ_A and σ_B are the variance of the images A and B of size $M\times N$ $C_1 = 6.5025,\ C_2 = 58.5225$ for 8 bit per pixel
6	UIQI	$UIQI = \frac{\sigma_{AB}}{\sigma_A\sigma_B}\cdot\frac{2\mu_A\mu_B}{\mu_A^2+\mu_B^2}\cdot\frac{2\sigma_A\sigma_B}{\sigma_A^2+\sigma_B^2}$	
7	SSIM	$SSIM = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2+\mu_B^2+C_1)(\sigma_A^2+\sigma_B^2+C_2)}$	
8	QDIFF	$(CC_{avg} - R) + SSIM$	R= RASE, if RASE is computed in percentage use R=RASE/100, CC_{avg} = CC/3, average cross correlation of R,G and B bands

Subjective analysis is carried out to compare the visual quality of the fused image, using 1000×1000 IKONOS images which are fused using the same six techniques used in quantitative analysis. Subjects are asked to rank the six images from 1 to 6 (1 is considered as best) based on visual quality (sharpness, color, brightness, contrast, overall quality); people are selected from variety of domains such as students, image processing experts, industry etc. The result of subjective analysis is shown in Table 3. From the table we can understand that most of the people ranked the proposed method as best and IHS as the bad image.

Table 2. Quantitative analysis of sample IKONOS and QuickBird images (1000×1000)

	IKONOS 1000x1000 image						QuickBird 1000x1000 image					
	IHS	Proposed	Chol	NSCT + IHS	Ada IHS	Imp IHS	IHS	Proposed	Chol	NSCT + IHS	Ada IHS	Imp IHS
RASE	45.2460	38.7459	51.4174	4.8126	65.0229	23.8175	26.9824	18.3354	10.2818	4.2225	7.8227	8.3906
ERGAS	11.5157	8.4085	9.3753	1.2025	15.5509	6.0714	6.7561	4.3637	2.5127	1.0570	1.9686	2.0988
SAM	3.0302	0.9945	3.0253	0.8287	2.6654	1.7284	1.5648	0.7269	0.9317	0.7029	0.7947	0.8977
UIQI Red	0.5120	0.6417	0.9491	0.5405	0.5508	0.7316	0.9588	0.9733	0.9554	0.9218	0.9171	0.9066
UIQI Green	0.6154	0.7378	0.9551	0.5963	0.6326	0.7511	0.9721	0.9750	0.9394	0.9148	0.9104	0.9019
UIQI Blue	0.4070	0.5233	0.8520	0.4166	0.4484	0.5507	0.9712	0.9693	0.9435	0.9122	0.9073	0.9012
CC Red	0.9808	0.8849	0.8883	0.3831	0.8933	0.4244	0.9938	0.9708	0.9412	0.8777	0.8848	0.8659
CC Green	0.9921	0.9394	0.9076	0.4244	0.8914	0.4834	0.9968	0.9822	0.9433	0.8752	0.8773	0.8624
CC Blue	0.9722	0.8000	0.8301	0.1566	0.8120	0.1887	0.9933	0.9683	0.9246	0.8465	0.8506	0.8352
SSIM	0.6856	0.7632	0.8199	0.4500	0.7058	0.5174	0.9716	0.9540	0.8795	0.7446	0.7802	0.7556
QDIFF	1.2148	1.2505	1.1811	0.7232	0.9212	0.6447	1.6964	1.7444	1.7131	1.5689	1.5728	1.5262
Time (sec)	0.2899	1.1126	0.1870	244.0583	1.5189	0.2647	0.1608	1.1074	0.2230	232.6192	1.1569	0.2031

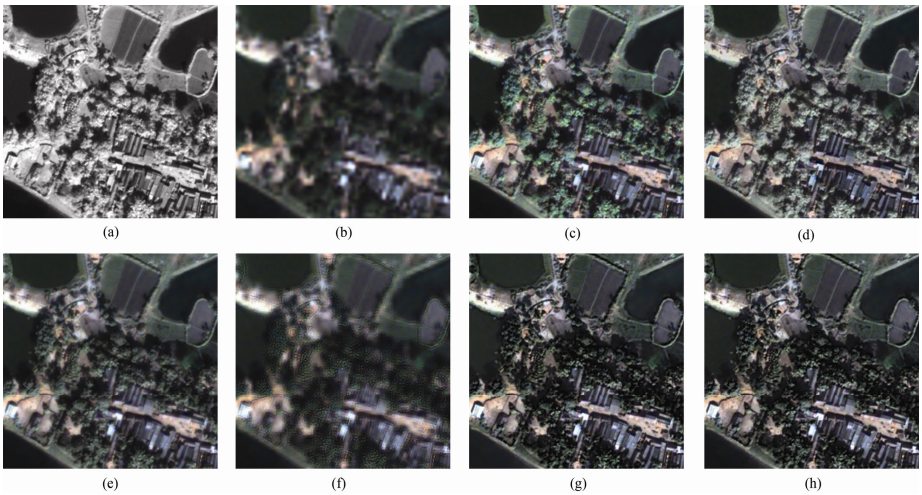


Fig. 1. QuickBird fused images (a) Panchromatic (b) Scaled up Multispectral RGB(c) Proposed (d) IHS (e) modified IHS by Choi (f) NSCT+ IHS (g) Improved-IHS (h) Adaptive IHS

Table 3. Subjective analysis results of sample IKONOS image (1000×1000)

	IHS	Proposed	Choi	NSCT + IHS	Ada IHS	Imp IHS
Avg Ranks	5.00	2.00	4.60	2.90	3.20	3.30

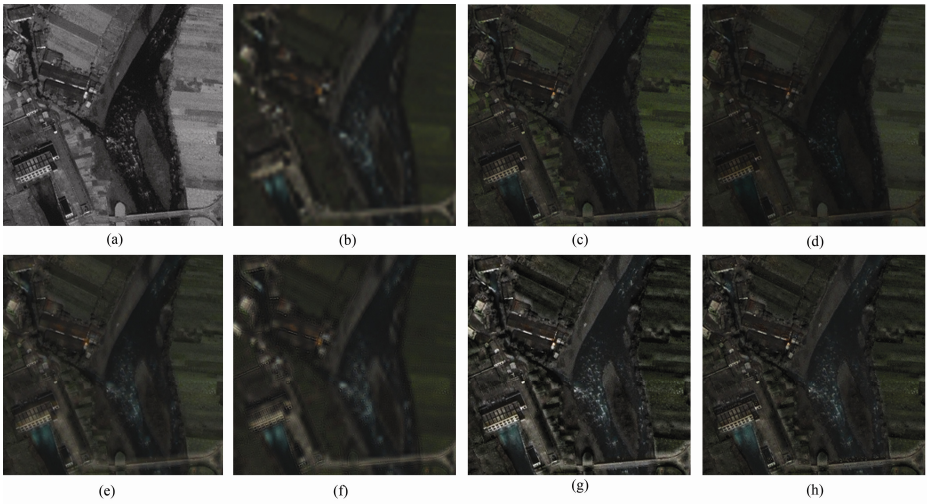


Fig. 2. IKONOS fused images (a) Panchromatic (b) Scaled up Multispectral RGB(c) Proposed (d) IHS (e) modified IHS by Choi (f) NSCT+ IHS (g) Improved-IHS (h) Adaptive IHS

4 Conclusion

In this work, a Bi-Level IHS image fusion algorithm for satellite images is proposed. From the quantitative and qualitative analysis, proposed method demonstrates good visual and

spatial qualities with reduced spectral distortion. Our method balances the spatial and spectral quality while improving the visual quality of the fused satellite images. The method is computationally fast and it can be implemented on a parallel computing architecture.

References

1. Haydan, R., Dalke, G.W., Henkel, J., Bare, J.E.: Applications of the IHS color transform to the processing of multisensor data and image enhancement. In: Proceedings of the International Symposium of Remote Sensing of Arid and Semi-arid Lands, Cairo, Egypt, pp. 599–616 (1982)
2. Eshtehari, A., Ebadi, H.: Image Fusion of Landsat ETM+ and Spot Satellite Images Using IHS, Brovey and PCA. Toosi Univ. Technol., Tehran (2008)
3. da Cunha, A.L., Zhou, J., Do, M.N.: The nonsubsampled contourlet transform: Theory, design, and applications. IEEE Trans. Image Process. 15(10), 3089–3101 (2006)
4. Shah, V.P., Younan, N.H., King, R.L.: An Efficient Pan-Sharpening Method via a Combined Adaptive PCA Approach and Contourlets. IEEE Transactions on Geoscience and Remote Sensing 46(5), 1323–1335 (2008)
5. Mahyari, A.G., Yazdi, M.: Panchromatic and Multispectral Image Fusion Based on Maximization of Both Spectral and Spatial Similarities. IEEE Transactions on Geoscience and Remote Sensing 49(6) (June 2011)
6. Li, D.: Fusion of Multispectral Remote Sensing Image and High Resolution Spatial Panchromatic image Based on NSCT and IHS. In: Second International Conference on Computer and Electrical Engineering (2009)
7. Kaplan, N.H., et al.: Fusion of multispectral and panchromatic images by combining bilateral filter and HIS transform. In: 20th European Signal Processing Conference (FUSIPCO 2012), Bucharest, Romania, August 27-31, pp. 2501–2505 (2012)
8. Rahmani, S., Strait, M., Merkurjev, D., Moeller, M., Wittman, T.: An Adaptive IHS Pan-Sharpening Method. IEEE Transactions on Geoscience and Remote Sensing 7(4), 746–750 (2010)
9. Choi, M., Kim, H., Cho, N.I., Kim, H.O.: An improved intensity-hue-saturation method for IKONOS image fusion. Korea Adv. Inst. Sci.Technol., Daejon, Korea. Tech. Rep. 06-9 (2008)
10. Choi, M.: A New Intensity-Hue-Saturation Fusion Approach to Image Fusion With a Tradeoff Parameter. IEEE Transactions on Geoscience and Remote Sensing 44(6), 1672–1682 (2006)
11. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Pearson Prentice Hall (2009)
12. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
13. Wald, L.: Data Fusion: Definitions and Architectures Fusion images of different spatial resolutions. Les Presses, EColedes Mines De Paris (2002)
14. Alperone, L., et al.: Comparison of pansharpening algorithms, Outcome of the 2006 GRS – S data Fusion contest. IEEE Trans. Geoscience and Remote Sensing 45(10), 3012–3021 (2007)
15. Wang, Z., Bovik, A.C.: A universal image quality index. IEEE Signal Process. Lett. 9(3), 81–84 (2002)

Visualisation of Multibeam Echosounder Measurement Data

Wojciech Maleika and Piotr Czapiewski

West Pomeranian University of Technology, Szczecin,
Faculty of Computer Science and Information Technology,
Zolnierska 52, 71-210, Szczecin, Poland
{wmaleika,pczapiewski}@wi.zut.edu.pl

Abstract. During a sea survey session performed using a multibeam echosounder millions of measurement points are generated. Sea surveys should be carried out in such a way, that the maximum accuracy of created seabed models (DTM) is achieved and the standards specified by the IHO S-44 guidelines are met. One of the requirements is so called full sea floor search, which means the ability of a system to detect all the cubic features at least 1 m in size. Spatial distribution of measurement points is irregular and the distances between closest data points are varying, depending on many factors (on survey parameters, depth or distance between the beam and the vessel). Due to those reasons, it is difficult for the users of hydrographic systems to evaluate the degree of coverage of seabed by measurement points, and therefore to confirm fulfilment of the normative requirements. As a solution we propose visualisation methods for measurement data collected in sea surveys. Specific features of such a visualisation are explained and a method for creating the images is presented, along with some exemplary results.

Keywords: data visualization, hydrographic systems, sea surveys.

1 Introduction

One of the most important yet also most difficult tasks involved in the process of building spatial information systems is creating a digital terrain model (DTM). This model is composed of the basic information layer used by GIS systems, and is used for either spatial mappings of parameters or assessments. Contemporary users of DTM are very demanding since they require a high quality of data (in terms of accuracy, reliability and validity), efficient data processing and visualization, and the ability to analyse the data in real time [3].

To create a DTM it is necessary to measure depths. Multibeam echosounders record depth observations, obtaining a large amount of data [5,10] – during one session, the depth and position of millions of points may be recorded. The cardinality and density of the collected data depends on the parameters of the recording device (i.e. the echosounder parameters: number of beams, sampling frequency and beam width; and the survey parameters: speed of the measurement vessel, number of survey tracks, and survey track configuration) [6,11].

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 373–380, 2013.

Based on the data collected during sea surveys, DTMs are created using appropriate processing algorithms [8,9]. The created DTMs and derived seabed charts must fulfil the norms specified by the International Hydrographical Organization [7]. The maing goal of this document is to specify the minimal hydrographic survey norms, to ensure the accuracy of the created models and maps. These norms also enable users to determine the spatial uncertainty of the data and effective utilization of them ensures the safe use of measurement results.

One of important parameters described in the regulations is the ability to perform so called full sea floor search. In practice, it is impossible to achieve 100% bathymetric coverage. Hence, a full sea floor search means a systematic method of exploring the sea floor to detect most of the cubic features of predefined size (1-2m for shallow waters). It should be noted that the IHO Special Order "feature detection requirements" of 1m and 2m cubes respectively, are minimum requirements. In certain circumstances it may be deemed necessary by the hydrographic offices to detect smaller features to minimise the risk of undetected hazards to surface navigation.

The evaluation of the number and distribution of measurement points, as well as the seabed coverage, is usually performed visually [2,10]. Contemporary hydrographic and GIS software allow for such a visualisation, but the image is usually created in a simplest way – for each measurement point a corresponding pixel is set in the output image (presented on screen) [11]. Since we are dealing with huge amount of data (many millions of points) with high density, usually the imaged area appears solid, without the possibility to distinguish particular points (see Fig. 1a). Only after a significant magnification of a fragment (see Fig. 1b) one can see precisely the number and distribution of points in a given subarea. Using such a solution it is however difficult to assess the distances between points, and hence to determine, if any cubic feature might remain undetected. Also, viewing larger areas in high magnification is cumbersome.

Given the above, it is worthwhile to develop a visualisation method for such data, allowing for a clear indication of places, where the DTM might not conform to IHO standards. The following paper presents two methods for visualising measurement data coming from multibeam echosounder, which aim to simplify the assessment of seabed coverage and to facilitate the detection of places, where the measurements might be insufficient to detect all the required features.

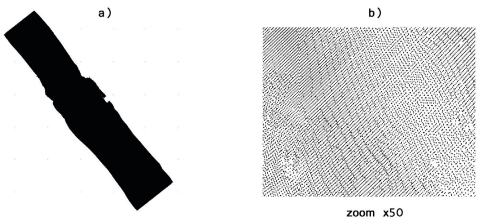


Fig. 1. Visualisation of measurement points in Surfer 9.0 software – whole area (a), a magnified fragment (b)

2 A New Method for Visualisation of Measurement Data Distribution

The measurement points coming from multibeam echosounder are placed in XYZ space (X and Y – position of measurement, Z – depth). XY values are given in one of reference systems (e.g. WGS84, EUREF-89, system 2000, UTM, etc.) [5]. In case of UTM system, the position of measurement is given in metric values, i.e. we note the distance on the X and Y axes from a selected base point. Data files saved in such a format simplify the calculation of distance between points (for other formats appropriate transformations must be performed).

Most hydrographic and GIS software allow for a 3D visualisation of measurement data. Unfortunately, the way the data is presented does not allow to assess the distribution or density of the points. There is also the possibility to present the data on a 2D image, but the imaging algorithms simply consist in determining for each measurement point a corresponding pixel and setting it to black. With such an approach, often multiple measurement points are mapped into the same image pixel. Furthermore, the size of created image is usually fitted to the software windows size and screen resolution, and does not depend on the size of visualised area. Such a visualisation of data makes it impossible to analyse the measurement data regarding the seabed coverage in a fast and precise way [1].

An important property, which should be considered , is a clear indication of not only the places, where the surveys were performed, but more importantly the places, where the density of measurement data is insufficient, which might result in some features remaining undetected.

Below are listed the basic features, that a visualisation method should possess, in order to allow for the assessment of the correctness of performed survey:

- possibility to define the size of areas or objects, that should always be detected (in metres),
- possibility to define the size of the output image,
- each subarea of a predefined size with no depth measurements performed must be clearly visible in the output image (the point must not be black out as a result of neighbouring points averaging or image downsampling).

The images created in that way will clearly show the places, where the density of data is insufficient and the IHO standard requirements might not be met.

Let us define the following variables:

- $grid$ – the cubic feature's size, which should always be detected,
- X_{size}, Y_{size} – the size of output image in pixels,
- X_{temp}, Y_{temp} – the size of intermediate image (its dimensions are adjusted to the area size and grid size),
- *src-file* – the name of a file with measurement data,
- x_{min}, x_{max}, y_{min}, y_{max} – edge values for measurement points (it is assumed, that this data is stored in metric UTM system),
- x_{data}, y_{data} – the coordinates of subsequent measurement points (UTM),
- x_{image}, y_{image} – subsequent pixel coordinates in the image.

In the presented method, the visualisation process consist of two basic stages. In the first stage, based on measurement data an intermediate high resolution image is created (adjusted to the area and grid size), where one pixel corresponds to a cubic feature size. In the second stage the final image (of size predefined by the user) is created, where the averaged information from the intermediate image is presented, with the addition of indication of areas with insufficient measurement points density.

The size of the intermediate image may be calculated using the formula:

$$\begin{cases} X_{temp} = \frac{X_{max} - X_{min}}{grid} \\ Y_{temp} = \frac{Y_{max} - Y_{min}}{grid} \end{cases} . \tag{1}$$

Next, for each measurement point the corresponding pixel coordinates in the intermediate image are determined, and the pixel is set:

$$\begin{cases} x_{image} = \frac{(x_{data} - X_{min})(X_{temp} - 1)}{X_{max} - X_{min}} + 1 \\ y_{image} = \frac{(y_{data} - Y_{min})(Y_{temp} - 1)}{Y_{max} - Y_{min}} + 1 \end{cases} . \tag{2}$$

After this stage the intermediate image contains a detailed map of measurement points distribution. Thanks to adjusting the image size to the size of area under investigation and to the grid size, blank pixels inside the area denote the subareas (the size of the grid) for which no measurements were taken. It could be interpreted as the possibility of missing in those subareas some features of a predefined size, hence the measurements do not comply with IHO regulations.

The dimensions of the intermediate image are usually bigger than the output image. For example, for an area 3x2 km in size and a grid size of 0.5 m the dimensions of intermediate image will be 6000x4000 pixels. The output image, however, should be displayed on screen and fit into software window, hence it will be definitely smaller. The simplest method of creating a smaller output image would be downsampling. Unfortunately, this would cause single white pixels present in the intermediate image to be averaged and not appear on the output image. Such a situation is undesirable, hence the scaling process should be performed in such way, that every white pixel in the intermediate image retains visibility in the output image. This approach will guarantee, that all areas with insufficient measurement data density will be visible in the output image.

The algorithm for creating images depicting spatial distribution of measurement points is presented in the Figure 2.

The results of the developed algorithm are presented using two real datasets from sea surveys, performed with Simrad EM3000 multibeam echosounder [4]. The parameters of test areas named "gate" and "rotator", as well as the dimensions of created intermediate images, are listed in Table 1.

In the Figure 3 a sample visualisation of measurement points distribution created using Surfer 8.0 software is shown (this is a typical visualisation performed using standard algorithms). Figure 4 presents a visualisation created using the method being described here. The images where created for 3 different grid sizes (1, 0.5 and 0.25 m). The output image has the size of 1300x1700 pixels.

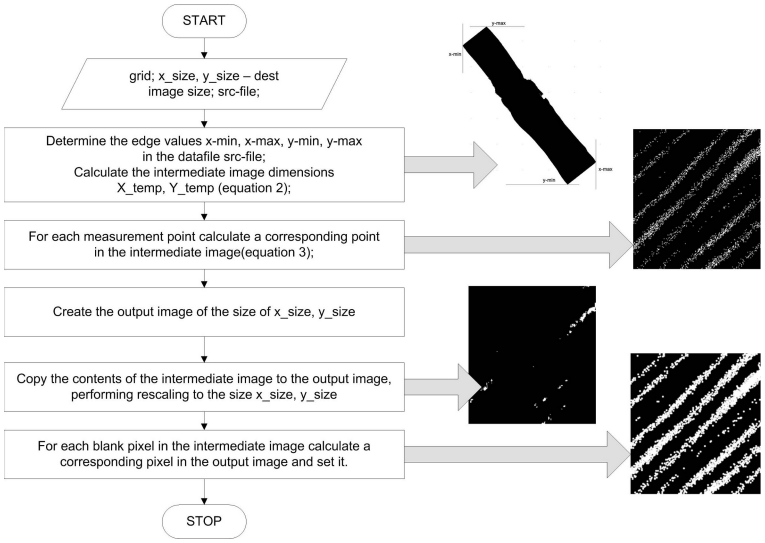


Fig. 2. The algorithm for creating images of measurement points spatial distribution

In case of commonly used today visualisation (see Fig. 3), the pixels denoting measurement points cover 100% of the area. The user can assume, that a full sea floor search had been achieved, which might not be the case. As we can see on images created using the proposed method, even for grid=1m there are some subareas, where the density of measurement points is too low (seen as dark spots in the image) and where some features located on the sea floor might remain undetected. The IHO regulations state, that most of the cubic features must be detected. It is up to the hydrographer to decide, whether the data density is sufficient or the occurring subareas of low density disqualify the survey.

If we want to make sure, that no objects bigger than 0.5 m were missed, the visualisation must be performed with grid set to 0.5 m. As we can see in the above images, the number of spots with insufficient density is higher, than for grid=1. In such case, the survey probably should not be considered conformant to the regulations. This effect is even more apparent for grid=0.25 m, where the number of blank spots is extremely high. Additional survey tracks should be performed, in order to gather more data.

A particular property of the images created this way should be noted. One should not assess the incorrect spots quantitatively (e.g. by estimating the percentage of area with insufficient data density). It is caused by the fact, that

Table 1. Properties of measurement data and created intermediate images

No.	Name	Number of meas. points	Area size [m]	Intermediate image size (grid=1m)	(grid=0.5m)	(grid=0.25m)
1.	Gate	3 812 445	811 × 951	811 × 951	1 622 × 1 902	3 244 × 3 804
2.	Rotator	12 263 261	1314 × 1848	1 314 × 1 848	2 628 × 3 696	5 256 × 7 392

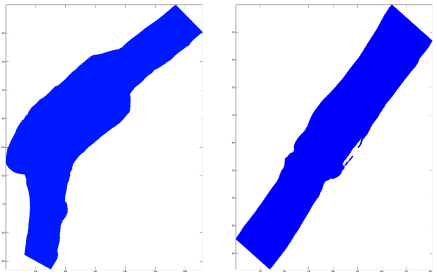


Fig. 3. Measurement points distribution visualisation created in Surfer 8.0 software

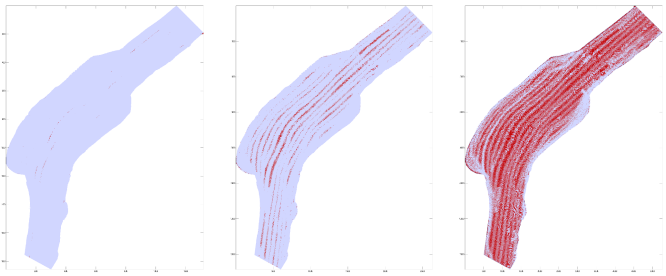


Fig. 4. Visualisation for the "rotator" surface. Grid size: 1m (a), 0.5m (b), 0.25m (c)

a single blank pixel in the intermediate image, denoting lack of measurement points in a very small subarea (e.g. 0.0625 m^2 for grid=0.25 m), in the output image remains blank, but corresponds to a much bigger area (1-4 m^2). In other words, each pixel in the output image corresponds to several pixels in the intermediate image (the number depends on the grid size and dimensions of the output image). If any of those pixels is blank, the pixel in the output image is also blank. This is why those images may be used only to show the locations, where data density is too low.

3 Quantitative Visualisation of Measurement Points Distribution

By performing the visualisation using the method described in the previous section, the user obtains only the information on the location of the spots, where the data density is insufficient. Therefore, from the informational point of view, we are dealing with binary images – yes/no, density sufficient/insufficient.

It seems desirable to prepare additional visualisations, which will allow to better assess the distribution of density of measurement points gathered in a given area. This should make it easier for hydrographic and GIS systems' users to perform quantitative analysis of data and to estimate the accuracy of created

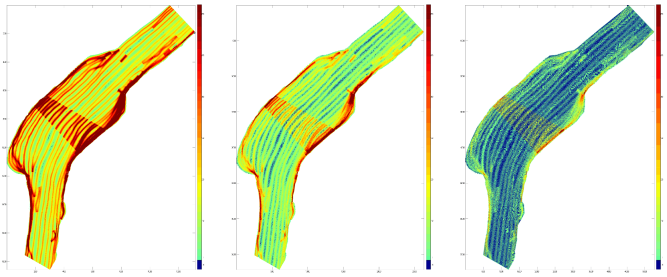


Fig. 5. Quantitative visualisation of the "rotator" surface. Grid size: 1m (a), 0.5m (b), 0.25m (c)

DTM models. It is obvious, that the bigger the amount of measurement data falling inside a single cell of the model, the lower the mapping error at this point. Quantitative estimation of measurement data allows to assess, whether it is possible to create a precise model, as well as to get additional information on the way the data were gathered.

The method for creating such a type of images is very similar to the one described in section 2. The only modification is making the colour of pixels in the output image dependent on the density of measurement points. In the intermediate image the value of a pixel is equal to the number of the measurement points related to it. In the next step, when determining the value of a pixel in the output image, the mean value of all corresponding pixels from intermediate image is considered. This approach allows for including in the output image the information on the average density of measurement points.

Examples of quantitative visualisation of measurement data are shown in the Figure 5.

In the above examples it is clearly visible, that for a grid size of 1 m, the density of the points near the edges of the areas is equal approximately 50-60 points/m^2. Inside the areas the density drops to approximately 10-40 points/m^2. The long red (dark) strips inside the areas show places, where the data from subsequent survey tracks overlap. Thanks to such type of visualisation the user is able to determine the degree of overlapping and the approximate layout of subsequent survey tracks. Also, the places with blank nodes, i.e. subareas with lacking data, are denoted (dark blue/black colour).

For smaller grid sizes (grid=0.5 m and grid=0.25 m) a significant decrease in density of measurement points is clearly visible. We can still recognize the distribution of points, layout of survey tracks and occurrences of blank nodes.

The grid size should be adjusted by the user to the size of the area under investigation, the expected accuracy of the model, IHO standards and to the characteristics of the area. For surfaces of low diversity or for typical surfaces, the grid size may be equal 1 m. For particular areas, such as those containing shipwrecks, underwater constructions, port terrains, a more precise grid of 0.5 m could be used.

4 Conclusions

Users of hydrographic systems do not possess any good tools allowing for visualisation of data coming from multibeam echosounders, so that the interpretation of this data could be easier. The method developed and described here allows for two novel ways of presenting the data, summarised below.

Visualisation of measurement points distribution shows the areas, where the density of measurement points is too low and it is possible to miss some cubic features of a predetermined size. Such type of images allows for quick analysis of measurement data regarding the conformance to IHO regulations.

Quantitative visualisation of measurement data allows for analysis of measurement points density distribution, estimation of survey tracks layout, and hence the assessment, whether the data is sufficient to create a precise DTM.

Institutions that may benefit from the conclusions drawn from the research are those that deal with conducting sea surveys, including bathymetric measurements. The results of the studies could prove useful to those fields in which the accuracy of seabed models is important, such as in oceanographic studies, geology, hydrotechnics, military science and tourism.

Novel imaging techniques might become a part of hydrographic and GIS software, therefore increasing the functionality scope of such software.

References

1. Bell, J.M., Linnett, L.M.: Simulation and Analysis of Synthetic Sidescan Sonar Images. IEE Proceedings, Radar, Sonar and Navigation (1997)
2. Calder, B.R., Mayer, L.A.: Automatic processing of high-rate, high-density multibeam echosounder data. Geochem. Geophys. Geosyst. 4(6), 1048–1069 (2003)
3. Hamilton, E.L.: Geoacoustic modeling of the sea floor. Journal of the Acoustical Society of America 68(5), 1313–1340 (1980)
4. EM3000 – Multibeam echo sounder. Kongsberg Maritime GmbH (November 2011), `http://www.kongsberg-simrad.de`
5. Etter, P.: Underwater Acoustic Modeling. Elsevier Applied Science (1991)
6. Gao, J.: Resolution and accuracy of terrain representation by grid DEMs at a microscale. Int. J. Geogr. Inf. Sci. 11(2), 199–212 (2001)
7. International Hydrographic Organization, IHO standards for hydrographic surveys, Publication No. 44, 4th edn. (1998)
8. Lubczonek, J., Stateczny, A.: Concept of neural model of the sea bottom surface. In: Rutkowski, L., Kacprzyk, J. (eds.) Neural Networks and Soft Computing, Advances in Soft Computing, pp. 861–866 (2003)
9. Maleika, W.: Development of a method for the estimation of multibeam echosounder measurement accuracy. Przeglad Elektrotechniczny 88(10b), 205 (2012)
10. Maleika, W., Pałczyński, M., Frejlichowski, D.: Multibeam echosounder simulator applying noise generator for the purpose of sea bottom visualisation. In: Maino, G., Foresti, G.L. (eds.) ICIAP 2011, Part II. LNCS, vol. 6979, pp. 285–293. Springer, Heidelberg (2011)
11. Yanalak, M.: Effect of GRIDding method on digital terrain model profile data based on scattered data. J. Comput. Civil. Eng. 17(1), 58–67 (2003)

Image Annotation in Presence of Noisy Labels

V. Chandrashekar[1], Shailesh Kumar[2], and C.V. Jawahar[1]

[1] IIIT-Hyderabad, India
[2] Google Inc., Hyderabad

Abstract. Labels associated with social images are valuable source of information for tasks of image annotation, understanding and retrieval. These labels are often found to be noisy, mainly due to the collaborative tagging activities of users. Existing methods on annotation have been developed and verified on noise free labels of images. In this paper, we propose a novel and generic framework that exploits the collective knowledge embedded in noisy label co-occurrence pairs to derive robust annotations. We compare our method with a well-known image annotation algorithm and show its superiority in terms of annotation accuracy on benchmark Corel5K and ESP datasets in presence of noisy labels.

Keywords: Image Annotation, Semantic concepts, Graph Mining.

1 Introduction and Related Work

Over the years, the Internet has become the largest database for multimedia content and is organized in a rich and complex way through tagging activities. One such example is collaborative tagging websites, such as Flickr, which collects millions of photos per month from tens of thousands of users. Consequently, there is immense research interest in producing efficient image annotation techniques for labelling social images to cope with the continuously growing amount of social image data.

Fig. 1. Figure shows example images and corresponding labels from both Corel5K datasets and Flickr images

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 381–389, 2013.

Existing annotation methods [1–4] consider the labels associated with the images to be devoid of errors and belonging to a small fixed vocabulary, and hence, can be directly used for designing annotation schemes. In contrast, the labels collected by collaborative tagging websites are noisy i.e misspelled, redundant, irrelevant to content, and unlimited in numbers. Thus, an interesting problem to address is, on how to use the noisy information available for annotating unlabelled images reliably.

Figure 1 shows examples of images from both expert-annotated Corel5K dataset and user tagged Flickr images. Labels of Flickr images like *love, life, emotions, excited etc.* are large in number and also irrelevant to the image content, whereas labels of Corel5K are often small in number and precisely describe the content of the image. In this paper, we address the problem of image annotation in presence of noisy labels.

Methods like WSABIE [3], which learn a low dimension embedding space for images and annotations, address this issue in an indirect way. Even Wordnet-based approach [5] has been used to remove irrelevant labels. In MLFDA [6], image annotation is posed as a multi-modal multi-class classification problem, where the noisy data is treated as a special kind of modal of the class and separating hyperplanes between classes are learned by kernel-based LFDA.

In this paper, we address the task of image annotation on noisy data using concept-modelling, a very popular notion in Information Retrieval community. The intuition is that, a specific meaning or aspect of an image can be well described by a group of highly related labels, referred to as label concept. Accordingly, each image can be organized into groups, each of which matches one label concept. This type of image organization not only removes noisy labels associated with an image, but also predicts additional labels that are actually depicted in the image but missing in the ground-truth annotations.

To show the utility of *concepts* over noisy systems, we compare its annotation performance with a baseline annotation method JEC [1], with noisy labels on Corel5K and ESP datasets. Our experimental results suggest that the proposed concept-based method leads to superior image annotation performance compared to JEC in presence of noisy labels.

2 Image Annotation in Presence of Noisy Labels

2.1 Nearest Neighbour Model for Annotation

K-nearest neighbour (or KNN) based methods [1, 2, 4] have been found to give some of the best results on the task of image annotation. The intuition behind them is that similar images share common labels. Most relevant KNN-based annotation methods are (i) JEC [1], which treats the annotation problem as retrieval and proposes a greedy algorithm for label transfer from neighbours, (ii) TagProp [4], a weighted KNN based method that transfers labels by taking weighted average of labels present among the neighbours, and (iii) 2PKNN [2], where a class-wise semantic neighbourhood is defined and only samples within this neighbourhood are used for annotation of unseen image. Since JEC is the

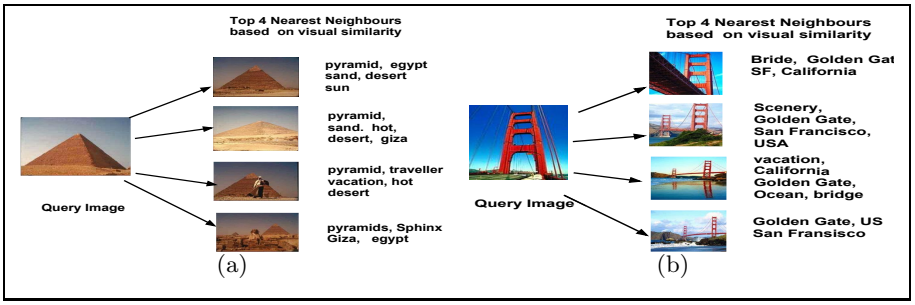


Fig. 2. Figure illustrates example of the well known K-Nearest Neighbour(K-NN) model used in image annotation. For each query image, the top 4 visually similar images are shown, along with the labels associated with them. The labels of the nearest neighbours are transferred to the query image for annotation.

essential backend method for modern successful techniques [2, 4], we compare our results with JEC [1] and show that our method is robust under noisy labels.

Let $I = \{ I_1, \ldots, I_N \}$ denote the collection of images and $V = \{ v_1, \ldots, v_m \}$ denote the *vocabulary* of m labels. The training set $T = \{ (I_1,V_1) \ldots (I_N,V_N) \}$ consists of pairs of images and their corresponding label sets, with each $V_i \subseteq V$. Given an unannotated image J, the task of annotation is to predict a set of labels that semantically describe J. In a typical NN-setting, we pick the top K visually similar images $T_J = \{ (T_{J,1},\gamma_{J,1}) \ldots (T_{J,K},\gamma_{J,K}) \}$. $\gamma_{J,K}$ denotes the visual similarity score of image J with its K^{th} neighbour, defined as:

$$\gamma_{J,K} = VisualSimilarity(I_J, T_{J,K}) \tag{1}$$

This score is generated as a function of distance between the images in visual feature space (SIFT, Color Histograms, GIST). Then, the labels of the nearest neighbours are ranked on basis of a label scoring function, κ_{J,v_i} and the top L labels are used to annotate the test image J. This label scoring function is usually based on frequency [1] or distance [4]. Figure 2 shows illustration of KNN model for image annotation.

2.2 Noisy Labels

In photo sharing websites, such as Flickr and Picasa, it is believed that most of the labels are correct, although there are many incorrect and redundant labels. Even from Figure 1, it can be observed that around 40-50% of the labels are irrelevant and are out-of context of the concept which the image represents.

Existing KNN-based methods [1, 2, 4] make an inherent assumption that labels present in the training set are reliable and correct, and hence can be directly used for training. They do not have an implicit mechanism of handling noisy labels and would not be suitable for annotation task in collaborative systems.

In this paper, we first present a graph-based approach for exemplifying the relationships between labels along with a noise removal algorithm to remove

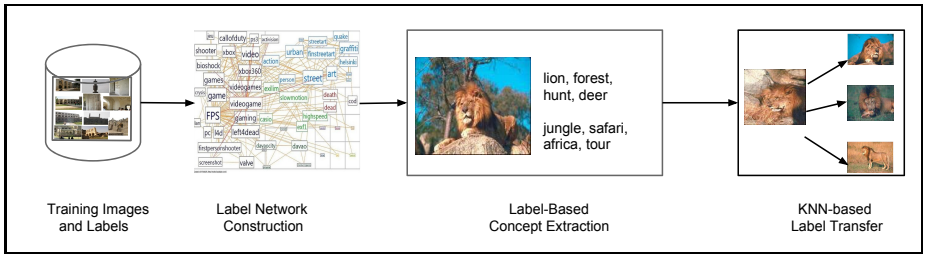


Fig. 3. An overview of our approach, which includes label network construction based on their co-occurrence, semantic concept identification using image labels and a KNN-based approach for transferring labels of concepts to unannotated image

most of the semantically-unrelated links among the labels. We then make use of this label network to infer the semantic concepts associated with images. Finally we illustrate how these concepts could be used for image annotation in a KNN-based setting. Figure 3 summarizes our approach.

2.3 Label Network Construction and Noise Removal

For label network creation, first the label co-occurrence counts, $\psi(\alpha, \beta)$ $\alpha, \beta \in V$, are calculated. But, this is not the best measure to quantify label co-occurrence strength as it may happen that two very frequent but uncorrelated tags might co-occur a lot compared to two relatively rare but correlated tags. Hence, we use *consistency*, $\phi(\alpha, \beta)$, to quantify associativity between labels, which is loosely defined as how much more likely is it to see the two labels together than random chance. We start by computing three types of raw statistics from the labels of training images: (i) Co-occurrence Counts $\psi(\alpha, \beta)$, (ii) Marginal Counts $\psi(\alpha)$, and (iii) Total Counts ψ_0 (defined in Step 3 of Algorithm 1). Joint probabilities $P(\alpha, \beta) = \frac{\psi(\alpha,\beta)}{\psi_0}$ and marginal probabilities $P(\alpha) = \frac{\psi(\alpha)}{\psi_0}$ are computed from these counts. These statistics are used for computing pair-wise consistencies between labels. We use *Normalized Point-Wise Mutual Information*[1], defined as

$$\phi(\alpha, \beta) = \frac{\log\left(\frac{P(\alpha,\beta)}{P(\alpha)P(\beta)}\right)}{-\log P(\alpha, \beta)} \forall \psi(\alpha, \beta) > 0 \tag{2}$$

to exemplify this consistency between labels, as it is a well-bounded quantity and suitably satisfies the definition of consistency.

Iterative Noise Removal: Initially, there is insufficient knowledge to identify which label-pairs are noise. After computing consistencies, label pairs with consistencies lower than a threshold θ_{consy} can be declared noise and are removed from the network. The marginal and total counts are then updated and consistencies are recomputed in the next iteration. The iterative noise removal method is described in Algorithm 1.

[1] `http://en.wikipedia.org/wiki/Pointwise_mutual_information`

Algorithm 1. DENOISE($[\psi(\alpha,\beta)]$)

Iteration $t \leftarrow 0$
$\psi^{(t)}(\alpha,\beta) \leftarrow \psi(\alpha,\beta)$
$\psi^{(t)}(\alpha) \leftarrow \sum_{\beta\in\mathbf{V}} \psi^{(t)}(\alpha,\beta)$, $\psi_0^{(t)} \leftarrow \frac{1}{2}\sum_{\alpha\in\mathbf{V}}\sum_{\beta\in\mathbf{V}} \psi^{(t)}(\alpha,\beta)$
while $\sum_{\alpha\in\mathbf{V}}\sum_{\beta\in\mathbf{V}} \phi^{(t)}(\alpha,\beta)$ converges **do**
  $\phi^{(t)}(\alpha,\beta) \leftarrow \text{CONSISTENCY}\left(\psi^{(t)}(\alpha,\beta), \psi^{(t)}(\alpha), \psi^{(t)}(\beta), \psi_0^{(t)}\right)$
  $\psi^{(t+1)}(\alpha,\beta) \leftarrow \psi^{(t)}(\alpha,\beta)\delta\left(\phi^{(t)}(\alpha,\beta) > \theta_{consy}\right)$ {$\delta(bool) = 1$ if *bool* is *true* else 0.}
  $\psi^{(t+1)}(\alpha) \leftarrow \sum_{\beta\in\mathbf{V}} \psi^{(t+1)}(\alpha,\beta)$, $\psi_0^{(t+1)} \leftarrow \frac{1}{2}\sum_{\alpha\in\mathbf{V}}\sum_{\beta\in\mathbf{V}} \psi^{(t+1)}(\alpha,\beta)$
  $t \leftarrow t+1$
end while

Table 1 shows effect of noise removal on the pair-wise consistencies of labels associated with label *water* in Corel5K data. It can be seen that consistencies of label *water* with correlated labels like *sea, ocean, beach, lake* increases significantly, whereas with irrelevant labels like *hills, grass* decreases to zero. By the end of this phase, we obtain a clean **noise-free** label network with pair-wise consistencies between labels as edge weights, which we will call *label consistency network*.

Table 1. Effect of noise removal on the consistencies of labels associated with label **water** in noisy Corel5K dataset, with $\theta_{consy} = 0.01$

Label	**sea**	**ocean**	**beach**	**lake**	**pool**	*hills*	*grass*
Before Denoising	0.3257	0.3720	0.3195	0.1699	0.0148	0.2081	0.1461
After Denoising	0.5750	0.5728	0.5658	0.3629	0.2449	0	0

2.4 Label-Based Concept Extraction

Here, we use the label consistency network for identifying semantic concepts associated with training images, using the image labels as seed. We define concepts as local maximal subgraphs in the label consistency network, based on a novel measure of *concept strength*, which is in-turn defined in terms of *label strength*.

In a systematic way, we first define *label strength* of a node (label) in a subgraph as a measure that captures the connectivity of the node with rest of the nodes in the subgraph. This essentially is the eigenvector centrality [7] of the subgraph. If $\mathbf{x} = \{x_1, x_2, ..., x_m\}$ be a set of m nodes in a subgraph and $\mathbf{W}(\mathbf{x}) = [\phi(x_i, x_j)]$ be the label consistency submatrix associated with this subgraph, then by eigenvector centrality, the *label strengths* converge to the first *unnormalized* eigenvector of $\mathbf{W}(\mathbf{x})$.

If $\lambda_1(\mathbf{W}(\mathbf{x}))$ is the first eigenvalue and $\mathbf{v}_1(\mathbf{W}(\mathbf{x}))$ is the first (normalized) eigenvector of this matrix, then *label strengths*, $\rho(\mathbf{x}|\mathbf{W}(\mathbf{x}))$, are defined by:

$$\rho(\mathbf{x}|\mathbf{W}(\mathbf{x})) = \lambda_1(\mathbf{W}(\mathbf{x})) \times \mathbf{v}_1(\mathbf{W}(\mathbf{x})) \tag{3}$$

$$\pi(\mathbf{x}|\Phi) = \min_{i=1...m} \{\rho_i\} \tag{4}$$

To capture the *tightness* of an arbitrary subgraph, we define *concept strength*, $\pi(\mathbf{x}|\Phi)$, to be **minimum** of the *label strengths* of all nodes (labels) of the subgraph (Equation 4). We now define *concepts*, as all those subgraphs in the label network, whose *concept strength* is higher than all its "neighbours". Neighbours of a subgraph, $\mathbf{x}$, is defined as all the subgraphs which can be obtained either by adding a single node ($\mathcal{N}_+(\mathbf{x})$) or removing a single node ($\mathcal{N}_-(\mathbf{x})$) from the given subgraph.

$$\mathcal{N}_+(\mathbf{x}) = \{\mathbf{y} = v \oplus \mathbf{x} | \forall v \in \mathbf{V} \backslash \mathbf{x}\} \qquad \mathcal{N}_-(\mathbf{x}) = \{\mathbf{y} = \mathbf{x} \backslash v | \forall v \in \mathbf{x}\} \tag{5}$$

We propose a greedy label-based approach to find such concepts, using the two atomic operations of **grow** and **shrink**. The **grow** operation tries to exhaustively find the best subgraph in $\mathcal{N}_+(\mathbf{x})$, which will have maximum *concept strength*, whereas the **shrink** operation finds the best subgraph in $\mathcal{N}_-(\mathbf{x})$.

Algorithm 2 explains how we extract multiple concepts associated with an image, using their labels as seed. The algorithm iterates over two phases: (i) **Shrink Phase**, which reduces labelset into a candidate subset of highly correlated labels, and (ii) **Grow Phase**, which adds more correlated labels to the candidate set making it a complete concept. The residue of the shrink phase is then again used as input over the next iteration to identify more concepts. Over this recursive process, multiple concepts associated with an image are identified.

2.5 Label Transfer for Annotation

Now, we illustrate how *concepts* can be used for the task of image annotation algorithm in an NN-setting. As a pre-processing step, we first use the training image labels to create a label consistency network and concepts associated with individual training images are extracted.

Given a test image J, we first find the top K-visually similar training images using features and metrics as suggested in [4]. Then, the labels associated with the concepts of the nearest training images are ranked based on a label scoring function, κ_{J,v_i}, defined as:

$$\kappa_{J,v_i} = \gamma_{J,K} \cdot \pi(\mathbf{x}|\Phi) \cdot \rho_{v_i}(\mathbf{x}) \tag{6}$$

This score is computed as product of visual similarity of the training image to the test image, concept strength of the concept associated with the training image and the label strength of the label within the concept. The individual components of the scoring function are first normalized before computing the scores. The labels are ranked based on this score and top L unique labels are assigned to the test image. Please note same label could have multiple scores, due to presence of same label in multiple concepts or images.

Algorithm 2. Label Based Concept Extraction($\mathbf{x}_0, \Phi$)

$\mathbf{x} \leftarrow \mathbf{x}_0$
$\mathbf{x}_{conc} = [\ ]$
while x **do**
  $[\mathbf{x}_{cand}, \mathbf{x}_{rem}]$ = ShrinkPhase($\mathbf{x}|\Phi$)
  $\mathbf{x}_{conc} = [\ \mathbf{x}_{conc}$ GrowPhase($\mathbf{x}_{cand}|\Phi$)] {Concepts extracted are concatenated}
  $\mathbf{x} \leftarrow \mathbf{x}_{rem}$
end while
A. **ShrinkPhase($\mathbf{x}|\Phi$)**
loop
  $[\mathbf{x}^-, \mathbf{x}_{rem}] \leftarrow Shrink(\mathbf{x}|\Phi)$ {Best possible *down-neighbor.*}
  if $\pi(\mathbf{x}^-) > \pi(\mathbf{x})$ **then**
    $\mathbf{x} \leftarrow \mathbf{x}^-$ {Not reached local maxima yet.}
  else
    return [x, $\mathbf{x}_{rem}$] {Reached local maxima.}
  end if
end loop
B. **GrowPhase($\mathbf{x}|\Phi$)**
$\mathbf{x}^+ \leftarrow Grow(\mathbf{x}|\Phi)$ {Best possible *up-neighbor.*}
while $\pi(\mathbf{x}^+) > \pi(\mathbf{x})$ **do**
  $\mathbf{x} \leftarrow \mathbf{x}^+$
  $\mathbf{x}^+ \leftarrow Grow(\mathbf{x}|\Phi)$
end while
return $\mathbf{x}^+$

3 Experiments

We present both qualitative and quantitative results, showing comparisons of our method with a very popular baseline method JEC [1], on benchmark annotation datasets: Corel5K and ESP [1].

As the motivation of our work is to show the effectiveness of our method on data with noisy labels, we create a parameter modulated noisy dataset by adding noisy labels to the training images of Corel5K and ESP. The noisy labels are taken from a vocabulary which has no overlap with the ground-truth vocabularies. We perform modulated experiments by regulating the degree of noise added to training data, using a parameter Q, which denotes the number of noisy labels added per training image. Annotation models are created using both, our method and JEC [1]. Evaluations are done using popular metric of mean F1-score over all the labels in the original vocabulary of the dataset. The F1-scores reported by our method correspond to label networks with threshold $\theta_{consy} = 0.01$, which was experimentally observed to be giving best results.

Figure 4 shows some qualitative results obtained using our method and JEC on noisy Corel5K data. It can be seen that some labels predicted by JEC are irrelevant and, also some ground truth annotations are missing in the predictions, whereas our method predicts all ground-truth annotations along with labels, which are depicted in the image but missing in the ground-truth annotation.

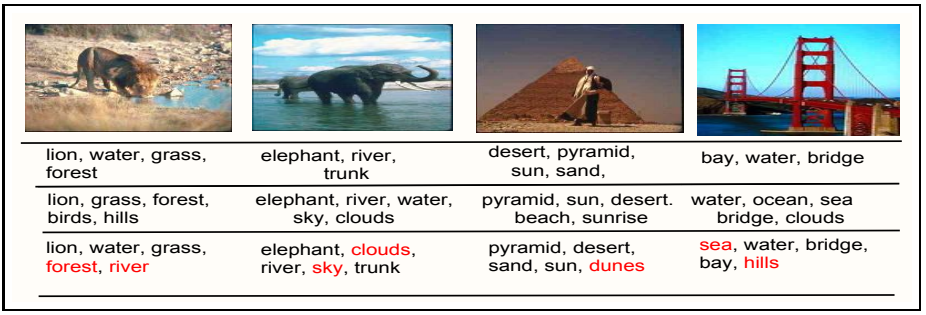
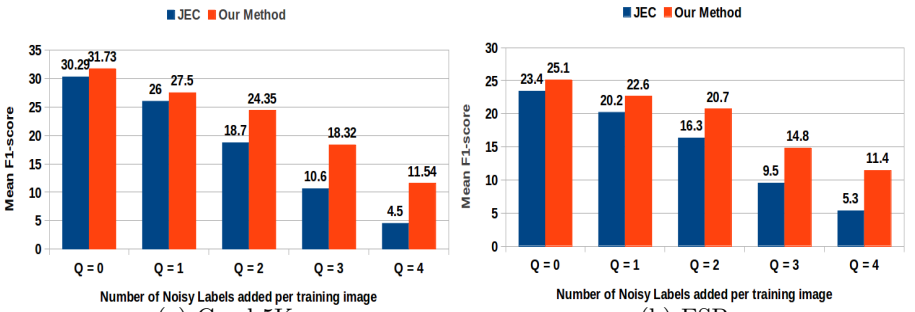


Fig. 4. Annotation of test images from noisy Corel5K dataset. The second, third and fourth rows show the ground truth labels, the labels predicted by JEC and the labels predicted by our method respectively. The labels in red are those, though depicted in the corresponding images, are missing in the ground-truth annotations and are predicted by our method.



(a) Corel-5K (b) ESP

Fig. 5. Comparison of annotation performance of our method and JEC[1] on noisy Corel5K and ESP datasets. Q denotes the number of noisy labels per training image.

To analyze our method's performance quantitatively, we compute the F1-score of each label in the ground-truth. The mean F1 scores using our method as well as those obtained by JEC [1] are reported in Figure 5. In both Corel5K and ESP datasets, as noise increases, F1-score of both methods decrease, but relatively our method performs better than JEC. In Corel5K, when only one noisy label is added per training image ($Q = 1$), there is about 6% improvement in F1-score. As Q is increased to 4, there is around 150% increase in the F1-score, which is a very significant improvement. This shows the effectiveness of using *concepts* in the task of image annotation, especially when noise is too high.

Experimentally we found that as θ_{consy} increases, the F1 scores also increase upto to a saturation point, and then start decreasing. This happens because once θ_{consy} reaches its saturation value, even relevant label-pairs in the network are considered as noise and discarded in the noise removal step. The pre-processing step of concept extraction takes considerable time. The label transfer step takes almost equal amount of time compared to JEC.

4 Conclusions

In this paper, we propose a novel knowledge-based approach for image annotation that exploits the semantic label concepts, derived based on the collective knowledge embedded in label co-occurrence based consistency network. An important future work to pursue is to build a *hierarchy of Concepts* and utilize them to learn useful insights for the tasks of image annotation and retrieval.

References

1. Makadia, A., Pavlovic, V., Kumar, S.: Baselines for image annotation. International Journal of Computer Vision 90(1), 88–105 (2010)
2. Verma, Y., Jawahar, C.V.: Image annotation using metric learning in semantic neighbourhoods. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 836–849. Springer, Heidelberg (2012)
3. Weston, J., Bengio, S., Usunier, N.: Wsabie: Scaling up to large vocabulary image annotation. In: IJCAI (2011)
4. Guillaumin, M., Mensink, T., Verbeek, J., Schmid, C.: Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In: ICCV, pp. 309–316 (2009)
5. Jin, Y., Khan, L., Wang, L., Awad, M.: Image annotations by combining multiple evidence and wordnet. ACM Multimedia (2005)
6. Wang, M., Zhou, X., Xu, H.: Web image annotation based on automatically obtained noisy training set. In: Zhang, Y., Yu, G., Bertino, E., Xu, G. (eds.) APWeb 2008. LNCS, vol. 4976, pp. 637–648. Springer, Heidelberg (2008)
7. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. Social Networks 32(3), 245–251 (2010)

Corpus Based Emotional Speech Synthesis in Hindi

Ravi Kalyan Bhakat, N.P. Narendra, and Krothapalli Sreenivasa Rao

Indian Institute of Technology Kharagpur, Kharagpur - 721302, West Bengal, India
{mailtome.rob,narendrasince1987}@gmail.com, ksrao@iitkgp.ac.in

Abstract. This paper explores a unit selection based concatenative approach towards emotional speech synthesis in Hindi. The emotions explored are sad and neutral. The Festival framework is used as the underlying Text-To-Speech (TTS) system. The various steps which are followed to create a new voice in Festival are described here. The developed TTS systems are evaluated by subjective evaluation tests. These tests indicate a significant improvement in the quality of synthesis after necessary prosody modifications. Finally, possible improvements which can be made on the systems are put forward.

Keywords: Emotional Speech Synthesis, Festival, Text to Speech, Unit Selection, Corpus based synthesis, Prosody modification.

1 Introduction

The rapid increase in the scale of human computer interaction in the recent past is evident for all to see. Starting from touch screens to bio metrics, gesture recognition to voice commands, the applications are incredible. Speech applications are a big part of this genre of computing. There has been considerable research on Text-To-Speech (TTS) synthesis. The speech synthesized by some TTS systems almost mimics human speech. Emotional Text-To-Speech (ETTS) synthesis however, is an area of great interest. The prosodic characteristics of speech, i.e. pitch, energy and duration vary with emotions. Incorporating these variations into speech while preserving its naturalness is a major issue. There are various methods that are used for ETTS synthesis. A rule based approach, which works with rules to change the voice parameters and impart emotions to speech was proposed in [1]. There are corpus based approaches which have a database of audio files and transcriptions of various emotions. The audio files are segmented into units and based on the text input, the required units are picked up and concatenated to form the speech [2] [3]. This approach gives us much more natural speech as compared to the rule based approach.

In this paper, we describe the development of corpus based TTS systems in Hindi which synthesize sad and neutral speech. They are syllable based speech synthesizers developed on the Festival TTS platform [4]. The baseline systems are improved by adding positional information to the syllables [2]. A fall-back system is incorporated to address the missing unit problem [2]. The systems are

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 390–395, 2013.

further improved by performing appropriate prosody modifications on the synthesized speech. The rest of the paper is organized as follows: Section 2 describes the speech corpora. Section 3 presents an overview of Festival and discusses the steps needed to create a new voice. Section 4 outlines some improvements made on the baseline system. Section 5 discusses the observations made on the systems. Section 6 discusses some possible improvements which can be made on the systems.

2 Details of the Emotional Speech Corpora

A sad and a neutral corpus of 1916 sentences each, serve as our database. The sentences were chosen from works of literature, story books and news articles. The corpora were selected based on maximum unit coverage by an optimal text selection algorithm [2]. The emotions reflected in the sentences were enacted by a professional female artist. The speech files were sampled at 16kHz and stored at 16bit PCM data format as .wav files. The corpora were cleaned by removing noisy and repeated files. The details of the corpora are given below:

1. **Neutral corpus:** Contains 1889 sentences with a duration of 2 hrs 54 mins and has a coverage of 4024 unique syllables.
2. **Sad corpus:** Contains 1915 sentences with a duration of 2 hrs 44 mins and has a coverage of 4107 unique syllables.

3 Building a Voice in Festival

The TTS systems are built on the Festival [5] framework using FestVox [6] and Speech Tools [7] as additional tools. The major issues addressed while building a new voice on Festival are enumerated below.

3.1 Selection of the Basic Unit

The selection of the basic unit in concatenative synthesis is an important decision. A proper choice results high quality speech with a reasonably sized corpus. Phones, syllables, words, phrases and sentences are some of the possible basic units. We have taken the syllable as the basic unit. In most Indian languages, as in Hindi, syllables correspond to written units. This makes the grapheme to phoneme conversion easier. A syllable is a string that can be generated by the regular expression C^*VC^*. Using this, syllables can be extracted from the input text. Figure 1 illustrates this concept with a sample utterance, *raajkumaari*, in Hindi. This consists of four syllables, *raaj*, *ku*, *maa* and *ri*. In the corpus, the speech files are segmented at the syllable level as illustrated.

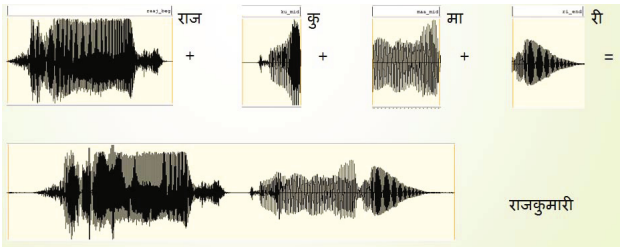


Fig. 1. Component syllables of the utterance *raajkumaari*

3.2 Labelling of the Corpus

The corpus needs to be labelled properly with accurate syllable boundaries for high quality synthesis. From the transcription, a phone level labelling is performed by an ergodic Hidden Markov Model [8] (eHMM) toolkit available on Festival. From the phone level segments, syllable level segments are extracted using syllabification rules. These labels are prone to alignment errors and need to be corrected to get exact syllable boundaries. The labels were manually adjusted to obtain proper syllable boundaries and reduce distortions in the synthesis.

3.3 Prosody Modelling

Prosody models are developed to predict the duration of the syllables using Classification and Regression Trees (CART). Phonetic, contextual and positional features are used in the process. Phonetic features include the positions of the various articulators in the vocal tract while the unit is spoken. Contextual features include the identity of the previous and the next unit. Positional features include the position of the unit in the word.

3.4 Building Unit Clusters

Each of the corpora have a coverage of around 4,100 syllables which are grouped together based on the acoustic differences. This helps in efficient retrieval of the target units. The clustering is done by *wagon* [5] by taking into account the phonological and positional features of the units. The given text input is first parsed and divided into syllables. The prosody models generate the target prosody of the units. From the unit clusters, a set of units is returned for each of the syllables based on the target cost. The unit selection module selects the best sequence of units from these sets which can be concatenated with the minimum possible distortion based on the concatenation cost. These costs can be optimized for better quality of synthesis [9] [10].

4 Improvements Made on the Baseline System

The prosody of the baseline syllable based TTS systems were not satisfactory. Firstly, this was because units from word beginnings were used to create word

ends and vice-versa. To address this, each of the syllables were tagged with their position such as *_beg*, *_mid* and *_end*. This ensured that positionally correct units were used during synthesis. Secondly, syllables absent in the database could not be synthesized. To address this, 500 sentences with phone level segmentation were kept in the corpus as a fall back. Thirdly, some neutral sentences had unsatisfactory prosody. The quality was improved by modifying the prosody with algorithms for pitch and duration modification [11] [12], speech normalization and energy correction. Some of these algorithms were integrated into the Festival framework for increased automation. A set of generic rules for prosody modification was generated by evaluating 300 neutral sentences and performing experiments to correct the observed distortions in the syllables. Table 1 outlines the rules which suggest the appropriate prosody modifications for the unsatisfactory units.

Table 1. Some rules generated to facilitate prosody modification of neutral sentences

Syllable	Problem noted	Modification required
me,ya,te,ti,ke,ki	higher pitch	pitch mod factor = 0.769
ne	higher pitch	pitch mod factor = 0.714
la,se	higher pitch	pitch mod factor = 0.833
kar,par,ab,le,de	lower pitch	pitch mod factor = 1.25
kha	lower pitch	pitch mod factor = 1.11
ka	short duration	duration mod factor = 1.3
ke	short duration	duration mod factor = 1.5
te,la	short duration	duration mod factor = 1.4
ge,pag,mag,te,re, ka,ya,le,me,tha,ne	loud	modify the intensity

5 System Evaluation and Observations

The evaluation of the systems was carried out in four phases. In the first phase, 500 sentences present in the corpus were synthesized by both the sad and neutral systems. Most of them were observed to be identical copies of the corpus files.

In the second phase, the emotional TTS systems were evaluated on intelligibility. A subjective evaluation was carried out with 15 subjects listening to 21 sentences synthesized by both the sad and neutral systems. All the subjects were between 21 to 30 years of age. Their Mean Opinion Score (MOS) were provided based on the scale given in Table 2. The neutral system was rated **3.388/5.00** and the sad system was rated **3.648/5.00**.

In the third phase, the emotion content in the sad sentences was evaluated. The subjects listened to the sad sentences and provided their scores based on the scale in Table 3. The sad system was rated **3.772/5.00**. Sentences which were not actually sad and simulated emotion received lower scores. On the other hand, inherently sad sentences received high scores indicating the natural integration of emotion during synthesis.

Table 2. Criteria for MOS scores on intelligibility of synthesized speech

MOS Score	Criteria
1	Very poor intelligibility, not understandable at all
2	Poor intelligibility, but some parts can be understood well
3	Average intelligibility
4	Good quality of speech, a few distortions
5	Very good quality of speech, can be understood easily

Table 3. Criteria for MOS scores on emotion content

MOS Score	Criteria
1	No emotion at all
2	Emotion evident in few parts of the sentence
3	Emotion present, but not perceived on first few hearings
4	Good quality of emotional speech, few unnatural parts
5	Very good quality of emotional speech, almost natural

From the above evaluations, it was noted that some neutral sentences had unwanted prosodic variations. Analysis revealed that there were some pitch and energy fluctuations that caused the sentences to sound unnatural. These issues were addressed by the prosody modification algorithms discussed in Section 4.

The fourth phase of system evaluation was conducted on the neutral system after the prosody modification algorithms were applied on the unsatisfactory sentences. A subjective evaluation was performed with 15 subjects listening to 10 modified sentences and providing their scores based on the scale in Table 2. They were rated **3.98/5.00** as compared to the earlier rating of 3.388/5.00. This indicates the effectiveness of the prosody modifications that were performed.

6 Summary and Future Work

In this work, systems which synthesize sad and neutral speech in Hindi were built on the Festival TTS system with syllables as basic units. Positional tagging and a fallback system system were used to improve the baseline TTS systems. Prosodically unsatisfactory sentences were modified using appropriate prosody modification algorithms to enhance their quality. Subjective evaluation of the systems were carried out and the results were discussed.

A better tagging approach with sentence level positional information may be explored for better synthesis. Selecting better units during synthesis can result in the lesser prosodic variations. This can be achieved by predicting the prosody of the target syllables in a better manner using some machine learning techniques. Incorporating the customized cost parameters into the TTS systems would help to increase the quality of synthesis. Better prosody modification techniques can be explored and may be useful in improving the synthesized speech.

The proposed prosody modification rule set can be extended to make it more comprehensive. Integrating those generic rules in the speech synthesis process will help increase the automation and quality of synthesis.

References

1. Murray, I.R., Arnott, J.L.: Implementation and testing of a system for producing emotion-by-rule in synthetic speech. Speech Communication 16(4), 369–390 (1995)
2. Narendra, N.P., Rao, K.S., Ghosh, K., Vempada, R.R., Maity, S.: Development of syllable-based text to speech synthesis system in Bengali. International Journal of Speech Technology 14(3), 167–181 (2011)
3. Iida, A., Campbell, N., Higuchi, F., Yasumura, M.: A corpus-based speech synthesis system with emotion. Speech Communication 40(12), 161–187 (2003)
4. Clark, A.J.R., Richmond, K., King, S.: Festival 2 - Build your own general purpose unit selection speech synthesiser. In: Proceedings of 5th ISCA Workshop on Speech Synthesis (2004)
5. Black, A.W., Taylor, P., Caley, R.: The Festival Speech Synthesis System, System documentation, edn. 1.4, for Festival Version 1.4.3 (2002)
6. Black, A.W., Lenzo, K.A.: Building Synthetic Voices. Language Technologies Institute, Carnegie Mellon University (2007)
7. King, S., Black, A.W., Taylor, P., Caley, R., Clark, R.: Edinburgh Speech Tools Library, System Documentation, edn. 1.2, for 1.2.3. Centre for Speech Technology, University of Edinburgh (2003)
8. Rabiner, L., Juang, B.H.: An introduction to hidden markov models. IEEE ASSP Magazine 3(1), 4–16 (1986)
9. Narendra, N.P., Rao, K.S.: Syllable specific unit selection cost functions for text-to-speech synthesis. ACM Transactions on Speech and Language Processing 9(3), 5:1–5:24 (2012)
10. Narendra, N.P., Rao, K.S.: Optimal weight tuning method for unit selection cost functions in syllable based text-to-speech synthesis. Applied Soft Computing 13(2), 773–781 (2013)
11. Rao, K.S., Yegnanarayana, B.: Prosody modification using instants of significant excitation. IEEE Transactions on Audio, Speech and Language Processing 14(3) (May 2006)
12. Rao, K.S., Prasanna, S.R.M., Yegnanarayana, B.: Determination of instants of significant excitation in speech using Hilbert envelope and group delay function. IEEE Signal Processing Letters 14(10) (October 2007)

Tracking Objects with Rigid Body Templates: An Iterative Constrained Linear Least Squares Approach

Satarupa Mukherjee[1], Nilanjan Ray[1], and Dipti Prasad Mukherjee[2]

[1] Department of Computing Science
University of Alberta, Edmonton, T6G2E8, AB, Canada
[2] Electronics and Communication Sciences Unit
Indian Statistical Institute, Kolkata
{satarupa,nray1}@ualberta.ca, dipti@isical.ac.in

Abstract. We formulate a novel iterative, constrained least squares technique for tracking rigid bodies. With barycentric representation of objects, we compute rigid body transformations under optical flow as iterative solutions of the optical flow constraint equation with a homogeneous, linear constraint. We show the efficacy of our method on cluttered videos.

Keywords: optical flow, constrained least squares, rigid body tracking.

1 Introduction

Rigid body tracking has many applications in image analysis. These include visual servoing of robotic arms, traffic monitoring and also medical imaging. There are different techniques available in the literature for rigid body tracking [3,8]. For rigid body tracking from a video, a popular approach is a two-step process [3,8]. The first step detects features or computes motion vectors. Various point features (e.g., corners), or structural features (e.g., edges) can be computed for this purpose. For motion vector computation, a common choice is optical flow. In the second step, a template is fitted to the features or motion vectors. A widely used mathematical tool for the second step is the linear least squares technique [8].

In this paper, we focus on rigid body tracking with the use of optical flow. There are different methods of optical flow computations available to date [2,9,10]. One significant disadvantage of these methods is that they lack prior motion information (e.g., rigid body motion). Thus, when used in tracking, these methods create a two-step pipeline, as mentioned earlier. The inherent difficulty with the family of two-step methods here is that parameters of the optical flow computation must be tuned properly to obtain meaningful tracking.

We propose a novel iterative and constrained linear least squares approach for tracking objects with rigid body motion. The principal and practical advantages of our method are that unlike the two-step tracking methods, there is

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 396–403, 2013.

no user tuning parameter in our technique and it is fast. Also, our method is able to handle partial occlusions because of the rigid body motion constraint. In our proposed method, the object template is represented in an affine coordinate system (also known as barycentric representation). Then, we solve optical flow constraint equation with rigid body constraints. The affine coordinate representation of the object allows us to map this problem to an iterative, constrained linear least squares approach. Thus, the process is a one-step process, and we never explicitly compute optical flow. The computation only involves solving linear equations with homogeneous linear constraints. So, standard least squares solvers are applicable here. As a challenging application, a publicly available surgical tool video [12] has been considered, where the surgical tool is to be tracked in a cluttered environment. The proposed method has been compared with another one-step optical flow based method using subspace constraint [7], a two-step optical flow method [2] and also with the exhaustive search method. It is observed that the proposed method and the exhaustive search method are most consistent in tracking the tool throughout the video sequences, whereas the other two methods failed to track the tool when it underwent sudden large motions. The record of the running times of all the methods show that the proposed method is an order of magnitude faster than its competitors. Our algorithm is also tested on six different vehicle videos having clutter and partial occlusions.

2 Optical Flow and Two-Step Tracking

The motion between two image frames is calculated using the optical flow method described as follows. A pixel at location (x, y, t) with intensity $I(x, y, t)$ is assumed to move by $\delta x, \delta y$ within a small time period δt between the two image frames. As the image intensity of the pixel is assumed to remain constant, it can be written as [4]:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t). \tag{1}$$

The movement of the pixel is assumed to be small. So, expanding the right hand side of (1) with Taylor series, we obtain:

$$I_x \dot{x} + I_y \dot{y} = -I_t, \tag{2}$$

where $\dot{x}$ and $\dot{y}$ are respectively the horizontal and the vertical velocities at pixel location (x, y); I_x, I_y and I_t are respectively the derivatives of the image in x, y and time directions. This is an equation with two unknown variables $\dot{x}$ and $\dot{y}$. Hence, it is an underdetermined system. This is known as the "aperture problem." Optical flow methods usually solve this problem by different regularization techniques, such as adding equations arising out of the assumptions about the smoothness of the flow.

The first step of a two-step tracking procedure computes optical flow in the region of interest typically using smoothness of the flow along with equation (2). We denote by (u_i, v_i) the displacement vector computed by optical flow at the

i^{th} pixel location. In the second step, rigid body motion of an object template is computed using the displacement vectors (u_i, v_i). Let (x_i, y_i) denote the i^{th} pixel coordinates in the object template. In order to compute the parameters of the rigid body motion, centering is applied first:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, p_i = x_i - \bar{x}, q_i = y_i - \bar{y}, \text{for } i = 1 \ldots N, \quad (3)$$

where N is the number of pixels that belong to the object template. Next, the rigid body transformation parameters a, b, c, d need to be solved from the following over-determined linear system:

$$\begin{bmatrix} p_i + u_i \\ q_i + v_i \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} p_i \\ q_i \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix}, \text{for } i = 1 \ldots N, \text{such that } a^2 + b^2 = 1. \quad (4)$$

First, we solve for c and d:

$$c = \frac{1}{N}\sum_{i=1}^{N} u_i, d = \frac{1}{N}\sum_{i=1}^{N} v_i. \quad (5)$$

From (4) and (5), we obtain :

$$\begin{bmatrix} p_i + u_i - c \\ q_i + v_i - d \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} p_i \\ q_i \end{bmatrix}, \text{such that } a^2 + b^2 = 1. \quad (6)$$

We solve for a and b from equation (6) with constrained least square minimization [6].

3 Proposed Method

Object tracking has been illustrated here as an application of affine combination [11]. The proposed tracking framework takes into account motion (affine/rigid body) and texture in the dense optical flow computation.

3.1 Affine Combination and Pixel Correspondence

If X (2-by-1 column vector) denotes a pixel location, then the 2D affine transformation of X is as follows: $Y = AX + f$, where A is a 2-by-2 matrix and f is a 2-by-1 column vector; A and f together represent a 2D affine transformation. X and Y are known as the corresponding point pair. For a set of points $X_1, X_2, X_3, ..., X_n$, a linear combination: $\sum_{i=1}^{n} \alpha_i X_i$ is called affine combination when scalar coefficients sum up to unity : $\sum_{i=1}^{n} \alpha_i = 1$ [11]. It is easy to establish that for corresponding point pairs (X_i, Y_i), the affine combinations $\sum_{i=1}^{n} \alpha_i X_i$ and $\sum_{i=1}^{n} \alpha_i Y_i$ are also corresponding points under the same affine transformation [11]. In other words, the affine combination parameters (α_i) are invariant under affine transformations. Since a rigid body transformation is a special type of affine transformation, the aforementioned invariance holds as well with a rigid body motion. We use this fact in our formulation for rigid body tracking.

3.2 Object Tracking Using Affine Combination

Let O denote the set of pixel locations belonging to the object to be tracked. This is obtained via segmentation (manual, semi-automatic, or automatic) on the initial video frame. Let us choose three corner points of the bounding box of O and denote them by X_1, X_2, and X_3. The set of parameters related with the affine combination of X_1, X_2, and X_3 is denoted by: $S = \{(s_1, s_2) : X_1 + s_1(X_2 - X_1) + s_2(X_3 - X_1) \in O\}$. Since the set S is invariant under rigid body transformations, tracking the object is equivalent to tracking the corner locations X_1, X_2, and X_3 over the image sequence. Notice that any three non-collinear points X_1, X_2, and X_3 can be utilized for our tracking framework. Let $(\dot{x}_k, \dot{y}_k), k = 1, 2, 3$, denote the velocity of the k^{th} corner X_k, then using (2), for any $(s_1, s_2) \in S$, we can write:

$$[(1 - s_1 - s_2)I_x \; s_1 I_x \; s_2 I_x \; (1 - s_1 - s_2)I_y \; s_1 I_y \; s_2 I_y][\dot{x}_1 \; \dot{x}_2 \; \dot{x}_3 \; \dot{y}_1 \; \dot{y}_2 \; \dot{y}_3]^T = - I_t, \forall (s_1, s_2) \in S. \quad (7)$$

Equation (7) has six unknown variables: $\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{y}_1, \dot{y}_2, \dot{y}_3$. Solutions are found by the least squares method. Note these variables only provide the velocities at the template corners. The corner coordinates are updated as follows. Let the k^{th} corner have coordinates, i.e., $X_k = (x_k, y_k)$. After solving (7), the corner coordinates are updated by a line search [5] in the direction of the velocities:

$$[x_1 \; x_2 \; x_3 \; y_1 \; y_2 \; y_3]^T \leftarrow [x_1 \; x_2 \; x_3 \; y_1 \; y_2 \; y_3]^T + a[\dot{x}_1 \; \dot{x}_2 \; \dot{x}_3 \; \dot{y}_1 \; \dot{y}_2 \; \dot{y}_3], \quad (8)$$

where

$$a = \underset{b \geq 0}{\text{argmin}} \sum_{(s_1, s_2) \in S} [I((1 - s_1 - s_2)(x_1 + b\dot{x}_1) + s_1(x_2 + b\dot{x}_2) + s_2(x_3 + b\dot{x}_3), (1 - s_1 - s_2)(y_1 + b\dot{y}_1) + s_1(y_2 + b\dot{y}_2) + s_2(y_3 + b\dot{y}_3)) - g(s_1, s_2)]^2. \quad (9)$$

$g(s_1, s_2)$ is a template function for the object. It is assigned the image intensity of the initial frame, i.e., the frame where tracking begins:

$$g(s_1, s_2) = I_0((1 - s_1 - s_2)x_1^0 + s_1 x_2^0 + s_2 x_3^0, (1 - s_1 - s_2)y_1^0 + s_1 y_2^0 + s_2 y_3^0). \quad (10)$$

The superscript 0 denotes the corners of the object on the initial frame I_0. The line search illustrated in (8)-(9) compares the object template intensity profile $g(s_1, s_2)$ with the current frame only along the direction of the computed velocity $[\dot{x}_1 \; \dot{x}_2 \; \dot{x}_3 \; \dot{y}_1 \; \dot{y}_2 \; \dot{y}_3]^T$. So, while the least squares technique computes the direction of displacement, the line search technique complements it with the magnitude of the displacement. Notice that because of the line search our technique is much less dependent on the multi-resolution pyramid computations that are almost always a strict requirement in the traditional optical flow computations.

3.3 Tracking with Rigid Body Constraints

If the object is assumed to have a rigid motion, then our aforementioned computational framework can accommodate such motions easily. The constraints imposed here are that the lengths between any two corner points for the object template remain constant,*i.e.*,

$$\begin{aligned}(x_1 - x_2)^2 + (y_1 - y_2)^2 &= C_1,\\ (x_2 - x_3)^2 + (y_2 - y_3)^2 &= C_2,\\ (x_3 - x_1)^2 + (y_3 - y_1)^2 &= C_3.\end{aligned} \tag{11}$$

When the above three equations are differentiated with respect to time and collected in the matrix vector form, the following equation is formed:

$$\begin{bmatrix}(x_1 - x_2) & (x_2 - x_1) & 0 & (y_1 - y_2) & (y_2 - y_1) & 0\\ 0 & (x_2 - x_3) & (x_3 - x_2) & 0 & (y_2 - y_3) & (y_3 - y_2)\\ (x_1 - x_3) & 0 & (x_3 - x_1) & (y_1 - y_3) & 0 & (y_3 - y_1)\end{bmatrix}$$

$$[\dot{x}_1\ \dot{x}_2\ \dot{x}_3\ \dot{y}_1\ \dot{y}_2\ \dot{y}_3]^T = \begin{bmatrix}0\\0\\0\end{bmatrix} \tag{12}$$

Thus, for a rigid body tracking, we solve (7) with the homogeneous linear constraint (12). This is our proposed constrained linear least squares solution for the rigid body tracking problem. Notice that for a rigid body motion constraint, at least three equations as in (12) are required.

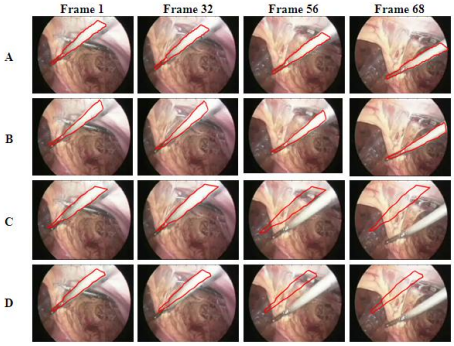


Fig. 1. Tracking results on tool sequence A. Proposed algorithm B. ES1 C. Irani's method [7] D. Brox optical flow method [2].

4 Results and Discussion

The proposed algorithm has been tested on 120 frames of a surgical tool video for tracking tool [12]. This video has significant amount of clutter, which makes

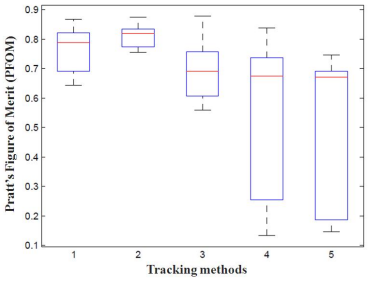


Fig. 2. Accuracy plot of four tracking methods : 1. Prpoposed Method 2. ES1 3. ES2 4. Irani's Method [7] 5. Brox et al's Method [2]

tracking a challenging task. Our algorithm is able to track the tool successfully in all the subsequent 120 frames in spite of severe clutter. Figure 1, row A shows various frames and our proposed rigid body tracker. For implementing the proposed algorithm, we choose Matlab's lsqlin function. To illustrate the accuracy of the algorithm Pratt's Figure of Merit (PFOM) [1] has been used and accuracies of segmentation for all the frames have been represented as a boxplot in Figure 2. The average time taken by our algorithms on an image frame with our Matlab implementation has been illustrated in Table 1, which shows that the proposed method is the fastest among its competitors.

The proposed method is compared with three other methods: exhaustive search method, a one-step method using subspace constraints [7] and a two-step method using optical flow computation by Brox et al.'s method [2]. Figure 1, rows B through D show the performances of these methods on various frames of the surgical tool video sequence. These competing methods are described here.

We use two different settings for the exhaustive search. The first setting uses unit spacing in horizontal, vertical and angular directions and is referred to as "ES1" method. It can be observed that although ES1 performs better in terms of accuracy Figure 2, it is about two orders of magnitude slower than our proposed algorithm (Table 1). So, it can be concluded that a better accuracy can be achieved by the exhaustive search algorithm only at the cost of running time. When we use a spacing of 4 for the horizontal, vertical and angular directions, the performance of tracking becomes inferior as Figure 2 shows. This method is referred to as "ES2." Note that the computation time of ES2 is still much inferior to the proposed method (Table 1).

The proposed method has been also compared with another well established optical flow based method [7], which has been developed on rigid body scenarios on the assumption that the set of all displacements of all points across several frames often resides in a low dimensional linear subspace. The main idea of this algorithm was to extend the two-frame Lucas-Kanade [10] region based constraint as well as the two-frame brightness constancy constraint to a generalized multi-frame method using the subspace constraints. Results show that our method is more competitive with respect to both accuracy and time (Figure 1, Figure 2, Table 1).

Table 1. Running time comparison

Tracking Algorithms	Time taken between two consecutive frames (s)
Proposed algorithm	0.3
ES1	17.79
ES2	4.52
Irani's method	9.52
Brox et al.	4.21

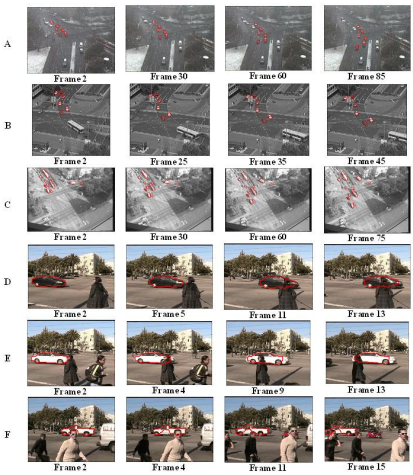


Fig. 3. Tracking results on vehicle sequences A. Car sequence in snowfall B. Car sequence (leftmost car having partial occlusion) C. Vehicle sequence having different types of vehicles D,E,F. Vehicles moving on a busy street occluded by pedestrians.

Finally, the proposed method has been compared with a two-step optical flow tracking [2], using rigid body constraints explained in Section 2. This optical flow method was implemented using multiresolution pyramid. This method needed some parameter tunings like the pyramid level for the multiresolution technique. In comparison, our algorithm did not need any parameter tuning. For the two-step algorithm, a 6 level pyramid with 1.25 resolution spacing has been considered. While for the proposed tracking method occurred at the original resolution. It is observed from Figure 2 that the two-step method has much inferior accuracy than the proposed method. Running times in Table 1 show that the two-step method is at least an order of magnitude slower than the proposed method.

In Figure 1, the tracking results of our algorithm with rigid body constraints have been compared with the results of the three other compared methods. It can be observed that from frame 56, when the tool underwent a large motion, Irani's method and the two step optical flow method failed to track it while our algorithm and the exhaustive search algorithm were able to keep track of it

correctly. A boxplot illustrating the PFOM values for all the four algorithms has been shown in Figure 2, which illustrates that the proposed algorithm is slightly less in accuracy than the exhaustive search, but it outperforms exhaustive search in running time which is shown in Table 1.

The visual results of tracking cars with the proposed method in different videos have been shown in Figure 3. The first row has shown the tracking results of a car sequence in a heavily cluttered video with snowfall. The second row has illustrated the tracking results of another car sequence where the left most car has been successfully tracked in spite of being partially occluded by a post on its way. The third row has demonstrated the tracking results of a sequence including different types of vehicles very small in size. Finally, fourth, fifth and sixth rows show the results from a publicly available dataset [13] where the vehicles are moving on a busy street and they are often occluded by pedestrians passing by. The video outputs and Matlab codes of our techniques are available here: `http://webdocs.cs.ualberta.ca/~satarupa/tracking.html`.

References

1. Abdou, I.E., Pratt, W.K.: Quantitative design and evaluation of enhancement/thresholding edge detectors. Proc. IEEE 67, 753–763 (1979)
2. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
3. Drummond, T., Cipolla, R.: Real-time visual tracking of complex structures. IEEE Trans. Pattern Anal. Machine Intell. 24, 932–946 (2002)
4. Fleet, D.J., Weiss, Y.: Optical Flow Estimation. In: Handbook of Mathematical Models in Computer Vision. Springer (2006)
5. Fletcher, R.: Practical Methods of Optimization. John Wiley and Sons, West Sussex (2006)
6. Golub, G.H., Van-Loan, C.F.: Constrained least squares. In: Matrix Computations. JHU Press, Baltimore (1996)
7. Irani, M.: Multi-Frame Correspondence Estimation Using Subspace Constraints. IJCV 48(3), 173–194 (2002)
8. Lipetit, V., Fua, T.: Monocular model based 3D tracking of rigid objects: a survey. Foundations & Trends in Comp. Graph. & Vis. 1(1), 1–89 (2005)
9. Horn, B.K.P., Schunck, B.G.: Determining optical flow. AI 17, 185–203 (1981)
10. Lucas, B.D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. IJCAI, 674–689 (1981)
11. Ray, N.: Image registration and object tracking via affine combination. In: ICAPR, pp. 175–179 (2007)
12. `http://www.dailymotion.com/video/x6if7d_less-price-ventral-hernia-surgery-i_tech`
13. `http://www.eetimes.com/design/military-aerospace-design/4017685/Background-subtraction-part-1-MATLAB-models`

Duration Modeling Using Multi-model Based on Positional Information

Vempada Ramu Reddy and Krothapalli Sreenivasa Rao

School of Information Technology,
Indian Institute of Technology Kharagpur
Kharagpur - 721302, West Bengal, India
ramu.csc@gmail.com, ksrao@iitkgp.ac.in

Abstract. This paper proposes prediction of syllable durations by developing multi-models using positional information. The proposed multi-model consists of four models used for predicting the durations of syllables. Among them, one of the models is used for predicting the durations of syllables present in mono-syllabic words, and the remaining three models are meant for predicting the durations of syllables present at initial, middle and final positions of polysyllabic words. In this study, (i) linguistic constraints represented by positional, contextual and phonological features and (ii) production constraints represented by articulatory features are used for predicting the duration patterns. Feed-forward Neural Networks (FFNN) are used for developing the duration models using above mentioned features. It was found, that the prediction accuracy is improved using multi-models compared to single duration model.

Keywords: Multi-models, Duration prediction, Prediction accuracy, Feed-forward neural networks, Linguistic and Production constraints.

1 Introduction

Duration plays a vital role in human speech communication. The sequence of syllable durations is defined as duration patterns. Variation in duration patterns provide naturalness to speech. Human hearing system is highly sensitive to the variations in duration patterns. Hence, while developing speech synthesis systems, acquisition and incorporation of the duration knowledge is very much essential [1].

In speech signal, the duration of each unit is dictated by the linguistic and production constraints of the unit [2] [3]. Duration models are developed using linguistic constraints represented by positional, contextual and phonological (PCP) features, and production constraints represented by articulatory (A) features. From here onwards the combination of features representing linguistic and production constraints is referred as PCPA features [1][4]. In this study, multi-model based approach is proposed for improving the prediction accuracy of durations of syllables. Multi-models are explored in this work based on

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 404–409, 2013.

positional aspects of syllables at the word level. The main reason for proposing multi-models for duration prediction is to avoid the bias towards mean during the prediction of syllable durations by single neural network. From the speech corpus, we have observed that durations of syllables has strong influence on their position in the word. Therefore, if we model the durations of syllables separately, based on their position in the word, the bias problem imposed by single neural network may be reduced. The implicit knowledge of duration is usually captured by using modeling techniques. Neural networks are used in this work to capture the underlying interactions that exist between input and output features [5] [3] [6].

The paper is organized as follows: The speech database used for modeling the syllable durations is presented in Section 2. Section 3 describes the features used for predicting the duration patterns. Performance of the proposed multi-model along with the single duration model using neural networks is explained in Section 4. Conclusions of this paper are presented in Section 5.

2 Speech Database

The text utterances of speech database used for this study are collected from Bengali Anandabazar newspaper, various text books and story books which covers wide range of domains. The collected text corpus covers 7762 declarative sentences with 4372 unique syllables and 22382 unique words. The text is recorded with a professional female artist in a noiseless chamber. The duration of total recorded speech is around 10 hrs. The speech signal was sampled at 16 kHz and represented as 16 bit numbers. The speech utterances are segmented and labeled into syllable-like units using ergodic hidden Markov models (EHMM). For every utterance a labeled file is maintained which consists of syllables of the utterance and their timing information. The percentage of different syllable structures present in the database are V(8.20%), VC(3.50%), VCC (0.20%), CV(50.41%), CVC(32.26%), CVCC(1.05%), CCV(2.50%), CCVC(1.77%) and CCCV(0.11%), where C is a consonant and V is a vowel. It is observed that durations of syllables in the database vary from 50 to 560 ms with mean and standard deviations 212.9 ms and 80.6 ms, respectively.

3 Features for Modeling the Durations

It is known that there exists some inherent relationship between linguistic and production constraints of speech to the duration variation patterns in speech [1] [7]. The linguistic and production constraints are represented by positional, contextual, phonological and articulatory (PCPA) features [8] [6] [9] . The positional features are further classified based on syllable position in a word and sentence, and word position in a sentence. The detailed list of linguistic and production constraints are given in Tables 1 and 2, respectively.

Table 1. List of factors affecting duration patterns of syllables, features representing the factors and number of nodes needed for neural network to represent the features

Factors	Features	# Nodes
Syllable position in the sentence	Position of syllable from beginning of the sentence Position of syllable from end of the sentence Number of syllables in the sentence	3
Syllable position in the word	Position of syllable from beginning of the word Position of syllable from end of the word Number of syllables in the word	3
Word position in the sentence	Position of word from beginning of the sentence Position of word from end of the sentence Number of words in the sentence	3
Syllable identity	Segments of the syllable (consonants and vowels)	4
Context of the syllable	Previous syllable Following syllable	4 4
Syllable nucleus	Number of segments before the nucleus Number of segments after the nucleus Number of segments in a syllable	3

Table 2. List of articulatory features

Features	Description
vlen	Length of the vowel in a syllable (short, long, dipthong and schwa).
vheight	Height of the vowel in a syllable (high, mid and low).
vfront	Frontness of the vowel in syllable (front, mid and back).
vrnd	Lip roundness (no rounding and rounding).
ctype	Type of consonant (stop, fricative, affricative, nasal, and liquid).
cplace	Place or position of the production of the consonant (labial, alveolar, palatal, labio-dental, dental and velar).
cvox	Whether consonant is voiced or unvoiced (voiced and unvoiced).
asp	Whether consonant is aspirated or unaspirated (aspirated and unaspirated).
nuk	Whether consonant with nukta or not nukta (withnukta and withoutnukta).
fph	Type of first phone in a syllable (vowel, voiced consonant, unvoiced consonant, nasal, semivowel, nukta and fricative).
lph	Type of last phone in a syllable (vowel, voiced consonant, unvoiced consonant, nasal, semivowel, nukta and fricative).

4 Proposed Multi-model Based Approach

In this study, multi-model is developed by dividing the syllables into four groups namely syllables in mono-syllabic words, syllables at initial, middle and final positions in polysyllabic words. Feed-forward neural networks (FFNN) are used in this work for developing multi-model. For each group of syllables, a separate model

is prepared using FFNN. Finally, the multi-model used in this work consists of 4 FFNNs developed by 4 groups of syllables mentioned above. After grouping the syllables, it is observed that syllables from mono-syllabic words represent only 5% of total syllables, and the remaining 95% is distributed as 35%, 25% and 35% syllables from initial, middle and final positions of the poly-syllabic words. The prediction performance of multi-model is compared with the single duration model developed using PCPA features. In this work, single duration model refers to single FFNN trained with PCPA features of all syllables together. The details of neural networks and the performance of multi-model are discussed in the following subsections.

4.1 Neural Networks

In this work, a four layer feed-forward neural networks (FFNN) [8] [10] with input layer, two hidden layers and output layer is used to model the durations of syllables. The structure of the four layer FFNN is shown in Figure 1. In this work, the data consists of 177820 syllables is used for modeling the durations of sequence of syllables. Different structures were explored to obtain the optimal four layer FFNN, by incrementally varying the hidden layer neurons in between 5 to 100. The details of neural network is given in Table 3.

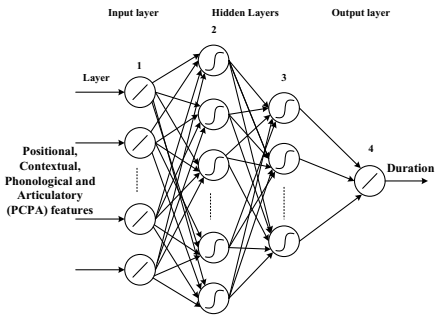


Fig. 1. Architecture of four layer feed-forward neural network

Table 3. Details of neural network

Non-linear activation function	tanh(s) where 's' is activation value
Training algorithm	Levenberg-Marquardt backpropagation
Train data	70%
Validation data	15%
Test data	15%

4.2 Evaluation

The prediction accuracy of the models is evaluated by means of objective measures such as average prediction error (μ), standard deviation (σ) and linear correlation coefficient ($\gamma_{X,Y}$). The computation of objective measures is given below:

$$D_i = \frac{|x_i - y_i|}{x_i} \times 100, \mu = \frac{\sum_i |x_i - y_i|}{N}, \sigma = \sqrt{\frac{\sum_i d_i^2}{N}}, \text{ and } \gamma_{X,Y} = \frac{V_{X,Y}}{\sigma_X . \sigma_Y}$$

$$\text{where } d_i = e_i - \mu, e_i = x_i - y_i, \text{ and } V_{X,Y} = \frac{\sum_i |x_i - \bar{x}| . |y_i - \bar{y}|}{N}$$

where x_i, y_i are the actual and predicted duration values, respectively, and e_i is the error between the actual and predicted duration values. The deviation in error is d_i, and N is the number of observed duration values of the syllables. σ_X, σ_Y are the standard deviations for the actual and predicted duration values respectively, and $V_{X,Y}$ is the correlation between the actual and predicted duration values.

The prediction performance of individual models of multi-model ($1-4$ rows), overall multi-model performance (5^{th} row) and performance of single model (6^{th} row) is given in Table 4. Column 1 of Table 4 shows type of model used for prediction. Columns 2-6 indicates the percentage of syllables predicted within different deviations from their actual duration values and columns 7-9 indicates the objective measures.

Table 4. Performance of FFNN based multi-model for predicting the duration values of syllables

Models	**% Predicted syllables within deviation**					**Objective measures**		
	2%	5%	10%	15%	25%	μ (ms)	σ (ms)	γ
Mono	8.33	19.02	38.36	56.71	77.22	55.46	44.18	0.83
Initial	10.39	23.85	42.07	62.20	81.62	26.59	22.22	0.89
Middle	7.49	18.31	31.25	50.16	73.66	29.60	24.74	0.83
Final	8.51	19.98	37.61	54.58	76.57	34.93	26.40	0.84
Multi	**8.90**	**20.86**	**37.60**	**56.23**	**77.63**	**31.71**	**25.42**	**0.86**
Single	7.96	18.88	35.14	50.63	72.56	39.04	35.09	0.83

From Table 4, it is observed that the performance of the model developed by using syllables representing initial position in word performed better than other models. From this we can hypothesize that initial position syllables are more discriminating than other syllables. It is also observed that the average prediction error of mono-syllables is high compared to others. The high average error of mono-syllables is mainly due to insufficient amount of syllables present in the database for training. However the overall performance of multi-model is outperformed compared to single duration model developed by using all syllables. From this hypothesis, we can conclude that the prediction accuracy of durations is improved by dividing the syllables and developing multi-model based

on the syllable position in the word. It was observed in the database that the mean durations of mono-syllables, initial, middle and final position of syllables in words vary greatly. Therefore, separating the syllables based on their position and developing the multi-model eliminated the biasses of one group of syllables towards mean values of other groups. This resulted in the improvement in the performance compared to single model.

5 Conclusions

In this work, prediction of durations of syllables is carried out using multi-model based approach. Multi-model based on syllable position in a word is developed by using neural networks. Linguistic and production constraints represented by positional, contextual, phonological and articulatory features are used for modeling the duration patterns. It is observed, that the performance of proposed multi-model based approach is performed better compared to single model. In future, prediction of durations can be analyzed by developing multi-models based on production aspects of speech segments.

References

1. Reddy, V.R., Rao, K.S.: Better human computer interaction by enhancing the quality of text-to-speech synthesis. In: Proc. Int. Conf. Intelligent Human Computer Interaction (IHCI), IIT Kharagpur, India, pp. 1–6 (December 2012)
2. Rao, K.S., Yegnanarayana, B.: Modeling durations of syllables using neural networks. Computer Speech and Language 21, 282–295 (2007)
3. Sreenivasa Rao, K., Mahadeva Prasanna, S.R., Yegnanarayana, B.: Two-stage duration model for Indian languages using neural networks. In: Pal, N.R., Kasabov, N., Mudi, R.K., Pal, S., Parui, S.K. (eds.) ICONIP 2004. LNCS, vol. 3316, pp. 1179–1185. Springer, Heidelberg (2004)
4. Reddy, V.R., Rao, K.S.: Intonation Modeling Using Linguistic, Production and Prosodic Constraints for Syllable based TTS Systems. Procedia Engineering, Elsevier 38, 2772–2783 (2012)
5. Yegnanarayana, B.: Artificial Neural Networks. Prentice-Hall, New Delhi (1999)
6. Reddy, V.R., Rao, K.S.: Intonation Modeling using FFNN for Syllable based Bengali Text To Speech Synthesis. In: Proc. Int. Conf. Computer and Communication Technology, MNNIT, Allahabad, pp. 334–339 (2011)
7. Rao, K.S., Yegnanarayana, B.: Intonation modeling for Indian languages. Computer Speech and Language 23, 240–256 (2009)
8. Reddy, V.R., Rao, K.S.: Two-Stage Intonation Modeling Using Feedforward Neural Networks for Syllable based Text-to-Speech Synthesis. Computer Speech and Language 27, 1105–1126 (2013)
9. Ramu Reddy, V., Sreenivasa Rao, K.: Intensity Modeling for Syllable Based Text-to-Speech Synthesis. In: Parashar, M., Kaushik, D., Rana, O.F., Samtaney, R., Yang, Y., Zomaya, A. (eds.) IC3 2012. CCIS, vol. 306, pp. 106–117. Springer, Heidelberg (2012)
10. Tamura, S., Tateishi, M.: Capabilities of a Four-Layered Feedforward Neural Network: Four Layers Versus Three. 8, 251–255 (1997)

Video Key Frame Extraction through Canonical Correlation Analysis and Graph Modularity

Rameswar Panda, Sanjay K. Kuanar, and Ananda S. Chowdhury

Department of Electronics and Telecommunication Engineering
Jadavpur University, Kolkata - 700032, India
{rameswar183,sanjay.kuanar}@gmail.com,
{aschowdhury@etce.jdvu.ac.in}

Abstract. Key frame based video summarization has emerged as an important area of multimedia research in recent times. In this paper, we propose a novel automated approach for video key frame extraction in compressed domain using canonical correlation analysis (CCA) and graph modularity. We prune certain edges from the Video Similarity Graph (VSG) using an iterative strategy until there is no improvement in graph modularity. Resulting connected components in the final VSG correspond to separate clusters. The proposed algorithm also uses multi-feature fusion using canonical correlation analysis to achieve higher semantic dependency between different video frames. Experimental results on some standard videos of different genre clearly indicate the superiority of the proposed method in terms of the F_1 measure.

Keywords: Key frames, Video Summarization, Canonical Correlation Analysis, Graph modularity.

1 Introduction

Recently, there has been a drastic increase in creation and storage of multimedia contents on the web. For example, as of now, one of the most primary video sharing web sites like YouTube reported that more than 1 billion unique users visit YouTube each month and 72 hours of video are uploaded every minute [13]. The evegrowing number of videos has motivated researchers to design efficient and effective video management schemes in order to provide a better overall multimedia experience to the consumer [1]. Summarization is mainly used to provide a condensed version of a full-length video through the identification of most important content within it [2,4]. Video skimming and key frame extraction are the two basic methods for summarizing videos [1-4]. A video of much shorter duration than the actual one is produced in the later case whereas the former deals with extracting some salient frames from the videos which preserves the overall content of a video with minimum data. Though the technique of skimming provides more information because of the audio and motion contents, key frames summarize the video content in a more rapid and compact manner. Moreover, key frame extraction can be used as a pre-processing step in various video analytics

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 410–416, 2013.

applications which suffer from the problem of processing large number of video frames [3]. The focus of this paper is extraction of key frames from the videos.

Different clustering techniques have been proposed in the literature to address the problem of extracting key frames from a video sequence [4-5, 10]. The performance of these methods heavily depends on user inputs and/or certain threshold parameters [5]. Moreover, many of those research works have focused on the uncompressed domain which makes the system unsuitable for online usage. Although the work presented in [6] is in compressed domain it utilizes the notion of similarity between successive frames. However, choice of similarity measures greatly influences the effective content representation of the key frame set. Most of the approaches use a single visual feature, *i.e.*, color to represent a video frame [5-6]. However, color alone cannot in general capture all the pictorial details needed to estimate the changes in the visual content of frames. Recently, Naveed et al. [3] proposed a key frame extraction technique based on a weighted combination of several features but their approach is too sensitive to selection of many control parameters.

In this paper, we present a novel approach for key frame extraction using multi-feature fusion via canonical correlation analysis (CCA) [7] and graph modularity clustering in compressed domain. Firstly, several features extracted from compressed domain I-frames are combined using CCA to form a single feature representing a video frame. Secondly, video similarity graph (VSG) is constructed over the video frames. Finally, an edge pruning strategy is repeated until there is no improvement in graph modularity.

2 Proposed Methodology

The proposed method consists of four main steps: (1) Feature extraction from compressed domain I-frames; (2) Multi-feature fusion with CCA; (3) Graph Modularity clustering; (4) Key frame extraction.

2.1 Feature Extraction

Since video data are usually available in compressed form, it is desirable to directly process the compressed video without decoding. Specifically, most video codecs (i.e., MPEG-1/2/4) are based on group of pictures(GOPs) as basic units [6]. The content of a GOP is represented using I-frames. The compression of the I-frames of a MPEG video is carried out by dividing the original image into 8 x 8 pixel blocks and transforming the pixels values of each block into 64 DCT coefficients [6]. Finally, on extracting the DC term of all the pixel blocks, a reduced version of the original image (known as DC image) is produced whose size is only 1/64 of original image.

In general, a single visual descriptor cannot capture all the pictorial details needed to estimate the changes in the visual content of frames. Hence we have extracted three features such as color (256 bins) [4], texture (i.e., edge, 80 bins) [4] and wavelet feature (20 bins) [9] from DC images.

2.2 Multi-feature Fusion

Since the feature extraction step produces three distinct feature vectors of different dimensions, we apply CCA to combine them [7]. In recent years, CCA has been applied to several fields as one of the most valuable multi-data processing methods [7] . CCA is a statistical method that finds a pair of directions which maximizes the correlation between projections of two random vectors. These projections are called canonical variates. This process of finding projections or directions is repeated until all the correlation features of the two random vectors are obtained. Algorithm 1 summarizes the steps involved in CCA fusion of the three features.

Algorithm 1. Multi-feature fusion
INPUT: Given three feature vectors: Color (C), Edge (E) and Wavelet (W)
OUTPUT: Canonically Correlated Feature Vector (CCFV)
1: **Procedure** (C, E, W, CCFV)
2: Assign X=C (256 elements), Y=E+W (100 elements).
3: Compute Covariance matrices S_{xx}, S_{yy} and between set covariance S_{xy}.
4: Compute G_1 and G_2.
$G_1 = S_{xx}^{-1/2} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1/2}$,
$G_2 = S_{yy}^{-1/2} S_{yx} S_{xx}^{-1} S_{xy} S_{yy}^{-1/2}$.
5: Compute orthogonal eigen vectors u_i, v_i and $r = rank\,(S_{xy})$
$\alpha_i = S_{xx}^{-1/2} u_i, \beta_i = S_{yy}^{-1/2} v_i, i = 1, 2, \cdots r$
6: Choose first $d\,(i.e., 100)$ pairs of α_i and β_i to make W_x and W_y.
$W_x = (\alpha_1, \alpha_2, \cdots, \alpha_d)\,, W_y = (\beta_1, \beta_2, \cdots, \beta_d)\,.$
7: Compute Canonically Correlated Feature Vector (CCFV)
$CCFV = {W_x}^T x + {W_y}^T y$
8: **End Procedure**.

2.3 Graph Modularity Clustering

We then construct the video similarity graph (VSG) using the data points (i.e., I-frames) in the refined feature space (CCFV) as its vertices. VSG is represented by $G = (V, E, W)$, where V is the set of nodes, E is the set of edges connecting the nodes and W is the set of weights corresponding to the strength of edges. The weights W_{ij} between two frames is defined as :-

$$W_{ij} = exp\left(-d_{ij}^2/\sigma^2\right) \tag{1}$$

where d_{ij} is the histogram intersection distance [9] between frames i and j. σ is a scaling parameter that determines the extent of similarity between two frames. As suggested in [8], $\sigma = \beta * max\,(d)$, where $\beta \leq 0.2$ and d is the set of all pairwise distances. In the VSG, the edges can be grouped into intra-cluster edges (edges whose end points are in the same cluster) and inter-cluster edges (edges whose end points are in different clusters). Our objective is to preserve

the intra-cluster edges and remove the inter-cluster edges which connect the individual clusters in an efficient manner. In our method, we prune certain edges depending on the difference between edge weight and scaling parameter, until there is no improvement in graph modularity [11]. Modularity$M(c_1, c_2, \cdots, c_k)$ of a graph clustering over k known clusters $c_1, c_2, \cdots, c_k$ is defined as

$$M(c_1, c_2, \cdots, c_k) = \sum_{i=1}^{k} \delta_{i,i} - \sum_{i \neq j} \delta_{i,j} \tag{2}$$

where $\delta_{i,j} = \sum_{\{v,u\} \in E, v \in c_i, u \in c_j} w(v,u)$, with each edge$\{v,u\} \in E$ included at most once in the computation. High value of modularity indicates good clustering. Remaining connected components of the final VSG after end of edge pruning represent individual clusters. Algorithm 2 summarizes the steps involved in graph modularity clustering.

Algorithm 2. Graph Modularity Clustering
INPUT: Video Similarity Graph (VSG), E, W, N= number of I-frames, σ
OUTPUT: $Clusters c_1, c_2, \cdots, c_k$
1: **Procedure** $(V, E, W, c_1, c_2, \cdots, c_k)$
2: *for* $i = 1$ *to* N
3: *for* $j = 1$ *to* N
4: $Dev_{ij} \stackrel{\Delta}{=} d_{ij} - \sigma$
5: End
6: End
7 : **Repeat**
8: Select edges which has high value of *Dev*, remove the edge from VSG
9: Find connected components from the VSG
10: Calculate Modularity (M)
11: **Until** no improvement in Modularity over two successive iterations.
12: Obtain individual clusters from final VSG.
13: **End Procedure**.

2.4 Key Frame Extraction

The frames which are closest to the centroids of each cluster are deemed as the key frames. Finally, the key frames are arranged in a temporal order to make the produced summary more understandable.

3 Experimental Results

3.1 Evaluation Dataset and Performance Measures

Ten video segments from Open Video (OV) projects [12] are used in the experiments to test the performance of our proposed method. The test set presents

a variety of video genres with different durations (46 sec. to 2 min). Each test video is in MPEG-1 format with a frame rate of 29.97 and the frames having dimensions of 352 x 240 pixels. Long videos are avoided due to limitation of annotation by a subject. The test set has also an intersection with the work in [5,6]. All the experiments are performed on a machine with Intel(R) core(TM) i5-2400 processor and 8 GB of DDR2-memory. Unlike other multimedia research areas, a consistent evaluation framework for video summarization is seriously missing possibly due to the lack of an objective ground-truth. In this work, we evaluated our technique based on the Comparison of User Summaries (CUS) mechanism proposed by Avila et al. [5]. This evaluation scheme is also adopted by various video summarization approaches for performance comparison [3, 5-6, 10]. In this evaluation scheme, the manually created user summaries are taken as references for comparison with the summaries generated by the automated methods. The performance was validated by F_1-measure which is a function of both precision and recall [6,10].

$$F_1 = \frac{2 * precision * recall}{precision * recall} \tag{3}$$

Here precision is the ratio of the number of matching frames to the total number of frames in the automatic summary and recall is the ratio of the number of matching frames to the total number of frames in the user summary.

3.2 Performance Comparison

In this section we make a comparative performance analysis of our proposed method with two well-known recent methods VSUMM [5] and VISON [6] using F_1-measure. Results of the proposed method can be seen at `https://sites.google.com/site/ivprgroup/result/vkcg`. The user summaries, and the key frames produced by the approach [5] are available at `https://www.sites.google.com/site/vsummsite/` and for the approach [6] are available at `http://www.liv.ic.unicamp.br/~jurandy/summaries/`. Table 1 presents the mean F_1-measure achieved by all the methods for several video categories. The results indicate that our proposed method performs better than VSUMM and VISON (improvement of 14.83% over VSUMM and 19.75% over VISON). In order to verify the statistical significance of those results, the confidence intervals for the differences between paired means are computed [5-6, 10]. Since the confidence intervals (with a confidence of 98%) do not include zero in both the cases (from Table 2), the results presented in Table 1 confirm that proposed method outperforms both VISON and VSUMM in statistically significant manner. Fig. 1 presents the key frames produced by the approaches for the video **The Voyage of the Lee, segment 15**. The user summaries for the same video are presented in Fig. 2. From Fig. 2, it can be noted that for most of the user summaries, our proposed method achieves higher F_1 value as compared to the other methods. The summary with the highest quality is achieved by our approach, which can also be confirmed by a visual comparison with the user summaries.

Table 1. Mean F_1 measure Comparative performance analysis for several video categories

Category	Videos	VSUMM	VISON	Proposed Method
Documentary	6	0.622	0.619	0.749
Educational	2	0.781	0.739	0.804
Lecture	2	0.758	0.673	0.860
Weighted Average	**10**	**0.681**	**0.653**	**0.782**

Table 2. Difference between Mean F_1 measure at a confidence of 98%

Method	Confidence Interval (98%)	
	Min.	Max.
Proposed method - VISON	0.142	0.269
Proposed method - VSUMM	0.094	0.173

Fig. 1. Summarization results for the video *The Voyage of the Lee, segment 15*: **Top row** →Proposed Method (Mean F_1-measure: 0.717), **Middle row** → VISON [6] (Mean F_1-measure: 0.515), **Bottom row** → VSUMM [5] (Mean F_1-measure: 0.666)

Fig. 2. User summaries for the video *The Voyage of the Lee, segment 15:* **Row1** → USER1: F_1(Proposed Method) = 0.591, F_1(VISON) = 0.362, F_1(VSUMM) = 0.568, **Row2** →USER2: F_1(Proposed Method) = 0.538, F_1(VISON) = 0.500, F_1(VSUMM) = 0.750, **Row3** → USER3: F_1(Proposed Method) = 0.768, F_1(VISON) = 0.428, F_1(VSUMM) = 0.600, **Row4** → USER4: F_1(Proposed Method) = 0.833, F_1(VISON) = 0.615, F_1(VSUMM) = 0.788, **Row5** → USER5: F_1(Proposed Method) = 0.857, F_1(VISON) = 0.671, F_1(VSUMM) = 0.626

4 Conclusion

In this paper, we propose a novel automatic key frame based video summarization technique using multi-feature fusion with canonical correlation analysis and graph modularity clustering in the compressed domain. Experimental results show that our technique outperforms the work described in [5] and [6]. Future work will focus on integration of visual attention model and inclusion of shape features to design an efficient system for search-and-retrieval of video sequences. Another direction of future research is to produce personalized key frames with various forms of unobtrusively sourced user-based informations.

References

1. Ejaz, N., Mehmood, I., Baik, S.W.: Efficient visual attention based framework for extracting key frames from videos. Signal Processing: Image Communication 28, 34–44 (2013)
2. Cong, Y., Yuan, J., Luo, J.: Towards Scalable Summarization of Consumer Videos Via Sparse Dictionary Selection. IEEE Transactions on Multimedia 14, 66–75 (2012)
3. Ejaz, N., Tariq, T.B., Baik, S.W.: Adaptive key frame extraction for video summarization using an aggregation mechanism. J. Visual Communication and Image Representation 23, 1031–1040 (2012)
4. Chowdhury, A.S., Kuanar, S.K., Panda, R., Das, M.N.: Video Storyboard Design using Delaunay Graphs. In: 21st IEEE International Conference on Pattern Recognition, pp. 3108–3111 (2012)
5. Avila, S.E.F., Lopes, A.P.B., Luz Jr., A., Araujo, A.A.: VSUMM: A mechanism designed to produce static video summaries and a novel evaluation method. Pattern Recognition Letters 32, 56–68 (2011)
6. Almeida, J., Leite, N.J., Torres, R.S.: VISON: Video Summarization for Online applications. Pattern Recognition Letters 33, 397–409 (2012)
7. Sun, Q.S., Zeng, S.G., Liu, Y., Heng, P.A., Xia, D.S.: A new method of feature fusion and its application in image recognition. Pattern Recognition Letters 38, 2437–2448 (2005)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
9. Ciocca, G., Schettini, R.: A innovative algorithm for key frame extraction in video summarization. J. of Real-time Image Processing 1, 69–88 (2006)
10. Panda, R., Kuanar, S.K., Chowdhury, A.S.: VISUC: Video Summarization With User Customization. In: IEEE International Conference on Communications, Devices and Intelligent Systems, pp. 89–92 (2012)
11. Schaeffer, S.E.: Graph clustering. Computer Science Review 1, 27–64 (2007)
12. The Open Video Project, `http://www.open-video.org`
13. YouTube Press Statistics, `http://www.youtube.com/t/press_statistics`

Semi-automated Magnification of Small Motions in Videos

Sushma M[1], Anubha Gupta[2], and Jayanthi Sivaswamy[1]

[1] Center for Visual Information Technology, IIIT Hyderabad, India
[2] Signal Processing and Communications Research Center, IIIT Hyderabad, India
sushma.m@research.iiit.ac.in,{agupta,jsivaswamy}@iiit.ac.in

Abstract. In this paper, we present a semi-automated method to magnify small motions in videos. This method amplifies invisible or hidden motions in videos. To achieve motion magnification, we process the spatial and temporal information obtained from the video itself. Advantage of this work is that it is application independent. Proposed technique estimates required parameters to get desirable results. We demonstrate performance on 9 different videos. Motion magnification performance is equivalent to existing manual methods.

Keywords: Motion Magnification, Time Frequency Analysis, Stockwell Transform, Automated Process.

1 Introduction

Human eye and its brain interface can visualize or detect the motion within a certain range of spatial and temporal frequencies. But in most of the cases, it might be possible that frequencies which are below this range also can have useful information. We can simplify this by saying that there can be small motions which are not visible to the naked eye. Even though these small motions are difficult to detect, they may contain useful information.

There are several applications for motion magnification, for example, visualization, physical diagnosis, pre-measurement planning for precise physical measurements, and surveillance etc. [7]. Therefore, there is a large scope for research to develop a system to amplify or magnify small motions.

There are previous works in this direction. In [1,2], human motions are generated by reusing the captured motion to create new motions. In this paper, we propose to process the data obtained from the video and reconstruct the video from the modified data such that new video shows magnified motion. In [3], small motion is magnified without tracking motion. However, this approach needs users to provide a set of parameters as input for every video in order to magnify motion. For unknown video, this consumes time to find desirable parameters. While all these techniques require user interaction in some or the other way, we are not aware of any previous work addressing automated approach to magnify small motions. This motivated us to develop a mechanism to estimate these parameters automatically from the given data. We use a time frequency representation

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 417–422, 2013.

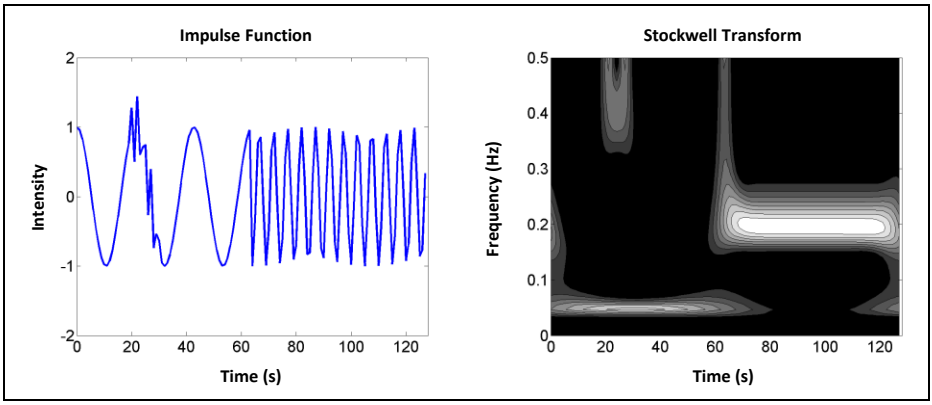


Fig. 1. An example of Stockwell transform. Time signal shown in left is $h[0:63] = \cos(2\pi t * 6/128), h[64:127] = \cos(2\pi t * 25/128), h[20:30] = h[20:30] + 0.5 * \cos(2\pi t * 52/128)$. It contains a low-frequency signal for the first half, a middle-frequency signal for the second half and a high-frequency burst at $t = 20$. All these frequencies are clearly visible along with time location in Stockwell transform as shown in right (bright pixels indicate high strength of transform).

called Stockwell transform for this purpose. We introduce computationally inexpensive techniques to estimate parameters. We illustrate the utility of the proposed method on examples in which small motions were made visible.

2 Background

2.1 Stockwell Transform

Even though Fourier transform gives the information about the spectral components in a signal, it fails to locate where those frequencies occur in that signal. So, it is preferable to consider time frequency representation (TFR). Different techniques for time frequency representation have been proposed. A few of them are short time Fourier transform (STFT), Gabor transform, continuous wavelet transform (CWT) and Wigner ville distribution etc. In [6], it was proven that Stockwell transform outperforms all these TFR techniques in localizing time and frequency because it has frequency dependent resolution whereas other transforms have windows of fixed width.

For a given time signal $h(t)$, its Stockwell transform is defined as,

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{-(\tau-t)^2 f^2/2} e^{-i2\pi f t} dt \tag{1}$$

where $h(t)$ is the time signal, f denotes the frequency and τ denotes time shift parameter. An example is shown for a synthetic signal in Fig. 1.

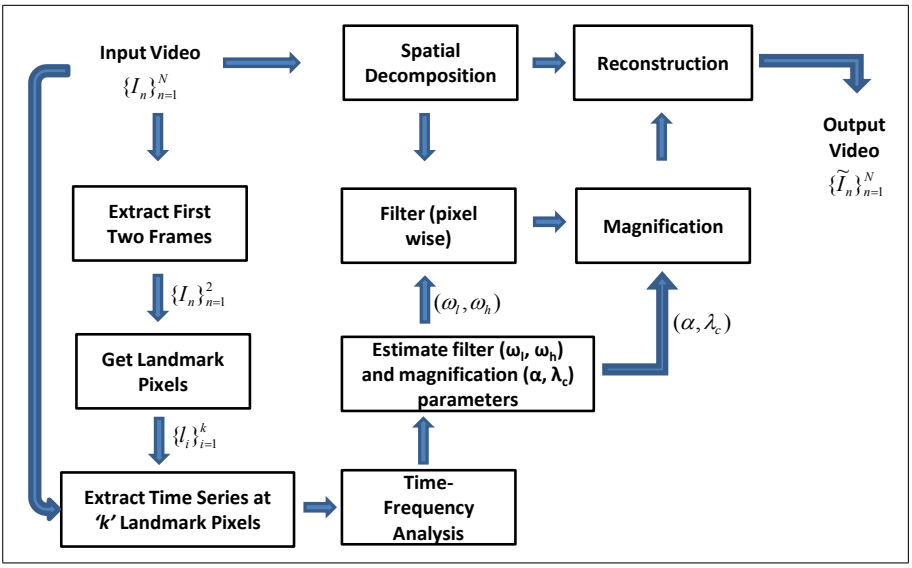


Fig. 2. Overview of the proposed framework. Initially, parameters (ω_l, ω_h, α, λ_c) are estimated from the landmark pixels ($\{l_i\}_{i=1}^{k}$) determined from the first two frames of input video ($\{I_n\}_{n=1}^{2}$). Then input video ($\{I_n\}_{n=1}^{N}$) is divided into spatial bands, which are pixel-wise filtered with same temporal filter and amplified by magnification factor, α. Reconstruction ($\{\tilde{I}_n\}_{n=1}^{N}$) is done by adding these amplified bands along with original spatial bands.

2.2 Semi-automated Video Magnification

Video magnification can be defined as magnifying or amplifying small motions in videos. This process helps in visualizing invisible or hidden deformations. It can be done by (i) detecting or tracking motion in a video and then by amplifying that motion [7,8]; (ii) directly magnifying motion without detecting or tracking [3]. We follow the second approach because tracking motion is computationally expensive. We magnify motion automatically unlike [3]. In [3], every video requires a set of parameters as input whereas in our method, we determine these parameters automatically by using the information from the video.

3 Parameter Estimation

To magnify small motions, in [3], given video is decomposed spatially. Then user has to set parameters for filtering and magnification. In our work, we estimate these parameters automatically. Main contribution of this work is to magnify the small motions in video automatically. It consists of the following two contributions: (1) Estimate bandwidth for temporal filter and (2) Estimate the magnification parameters, automatically.

Overview of the proposed framework is given in Fig. 2. Given an input video of N frames, $\{I_n\}_{n=1}^{N}$, the proposed method consists of the following steps: **(1) Estimation of parameters:** (i) Extract first two frames, $\{I_n\}_{n=1}^{2}$; (ii) Estimate landmark pixels, $\{l_i\}_{i=1}^{k}$; (iii) Extract time series at these 'k' pixels from N frames; (iv) Apply time frequency representation; (v) Estimate parameters for bandpass filter (ω_l, ω_h) and magnification (α, λ_c); **(2) Magnification:** (i) Decompose the given video spatially into different spatial bands; (ii) Apply pixel-wise bandpass filter with ω_l and ω_h as lower and higher cut-off frequencies; (iii) Multiply with magnification factor, α for wavelengths less than λ_c; **(3) Reconstruction:** (i) Reconstruct the video, $\{\tilde{I}_n\}_{n=1}^{N}$ by adding magnified signal to original signal.

To get the landmark pixels, it was observed that considering all pixels for determining bandwidth and magnification factor is not efficient due to the fact that (1) the whole process will be time consuming and (2) videos, considered for this work, have small motions in few regions and hence, all pixel locations do not necessarily undergo motion. Therefore, we adapted a mechanism to find landmark pixels. These pixels are obtained from the difference of edge maps of first two frames of a given video because the pixels at edges definitely experience motion from one frame to another frame. Time signals at these landmark pixels are used to determine the parameters.

Previous works, for example [5] has shown that temporal processing is done generally to extract invisible information from the signal. In general, spectral band is determined either empirically or assumed depending on application but such approaches fail to find the dominant frequency components automatically. We eliminate such dependency. In this work, we consider time signals from a video sequence, only at the locations of landmark pixels. We find bandwidth for bandpass filtering the time signals automatically. These limits of frequency are determined using Stockwell transform. It is observed that the minimum and maximum frequencies of mean of frequencies obtained from each time series of a pixel can give useful information for temporal processing. We carried out this processing on time series data in Y space of YIQ colour space.

From Stockwell transform, we estimate the magnification factor α. It is determined as follows,

$$(1 + \alpha)h(t) < \frac{\lambda_c}{8} \tag{2}$$

where $h(t)$ is time signal and λ_c is cut-off wavelength beyond which magnification factor, α is zero. Eq. (2) provides the largest magnification factor. As explained in Section 2.1, Stockwell transform gives time frequency representation. We utilize this time frequency representation to find desirable magnification factor α. From Stockwell transform, we obtain information about which wavelengths (or frequencies) are occurring at what times. Since videos involve small motion, we need to consider landmark pixels at which time signals can be used to determine magnification factor. A few of these time signals may contain noise and hence, may give incorrect α and λ_c values. To overcome this problem, we have considered median of all magnification factors and maximum of all wavelengths obtained from Eq. (2).

Spatial decomposition of videos into different spatial frequency bands is done to increase the signal-to- noise ratio using Laplacian pyramid [9]. These frequency bands are filtered and magnified pixel-wise differently according to the level in pyramid [3] using estimated parameters. Reconstruction is done by adding original signal from spatial decomposition and magnified signal.

4 Experiments and Results

We use data given in [4] for our experiments. This data consists of 9 videos, namely *baby*, *baby2*, *camera*, *face*, *face2*, *guitar*, *shadow*, *subway* and *wrist* as mentioned in Table 1. Sample frames from these videos are shown in Fig. 3. All experiments are implemented using MATLAB on a system with 4GB RAM and Intel® core i5 CPU with 2.5 GHz processor. Every video takes time in the order of a few minutes to compute.

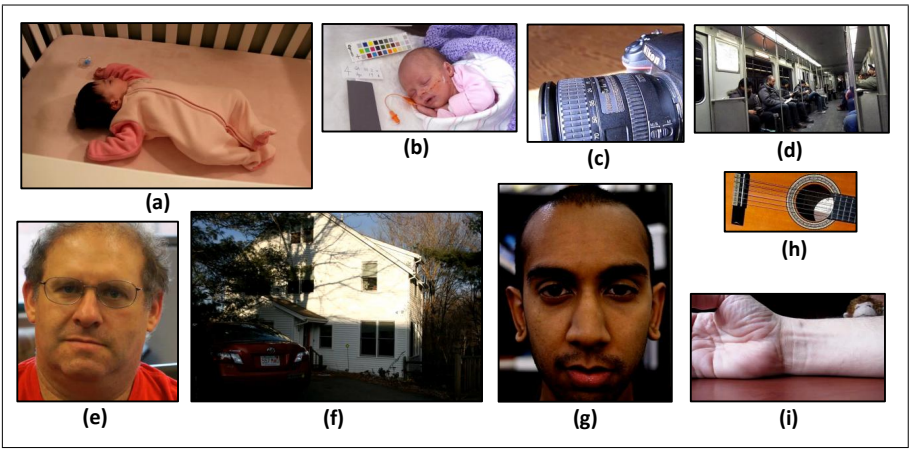


Fig. 3. Sample frames from videos used in experimentation (a)*baby* (b)*baby2* (c)*camera* (d)*subway* (e)*face* (f)*shadow* (g)*face2* (h)*guitar* (i)*wrist*. All these are shown in proportion with the size of the corresponding video.

Filter parameters, magnification factor and cut-off wavelength are determined from Stockwell transform by considering only landmark pixels. Obtained magnification factors α and cut-off wavelengths λ_c for each video are listed in Table 1. Even though for some videos, magnification factors are not close to the values mentioned in [3], reconstructed videos are comparable to those of [3]. These results are available on `http://researchweb.iiit.ac.in/~sushma.m/premiResults`.

Table 1. Estimated parameters for magnification factor α, cut-off wavelength λ_c, lower cut-off frequency for temporal filter ω_h, higher cut-off frequency for temporal filter ω_h.

Video	α		λ_c	ω_l(Hz)	ω_h(Hz)
	Proposed	[3]			
baby	32	10	30	0.5	3
baby2	30	150	30	0.6	2.8
camera	7	120	29	1.5	16.9
face	55	100	30	0.6	5.6
face2	17	20	30	1.02	10.56
guitar	101	100	29	0.77	8.16
shadow	13	5	30	0.92	5.8
subway	42	60	29	0.52	4.05
wrist	11	10	30	0.5	4

5 Conclusions and Future Work

This paper presents a semi-automated method for magnifying small motions in videos. We have demonstrated that the parameters required to magnify the motion can be generated in automated fashion by using the time frequency representation called Stockwell transform. However, it is observed that noise is introduced in some reconstructed videos. In future, we would modify the proposed method to reduce this noise and attempt to make this process fully automatic. We believe that this method can have potential applications in medical imaging.

References

1. Arikan, O., Forsyth, D.A.: Synthesizing constrained motions from examples. ACM Trans. on Graph (2002)
2. Lee, J., Chai, J., Reitsma, P.S.A., Hodgins, J.K., Pollard, N.S.: Interactive control of avatars animated with human motion data. ACM Trans. on Graph (2002)
3. Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J.V., Durand, F., Freeman, W.T.: Eulerian video magnification for revealing subtle changes in the world. SIGGRAPH (2012)
4. Eulerian Video Magnification for Revealing Subtle Changes in the World, `http://people.csail.mit.edu/mrub/vidmag/`
5. Poh, M.Z., Mcduff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optical Express (2010)
6. Stockwell, R.G., Manisha, L., Lowe, R.P.: Localization of the Complex Spectrum: The S Transform. IEEE Trans. on Signal Processing (1996)
7. Liu, C., Torralba, A., Freeman, W.T., Durand, F., Adelson, E.H.: Motion magnification. ACM Trans. Graph (2005)
8. Wang, J., Drucker, S.M., Agrawala, M., Cohen, M.F.: The cartoon animation filter. ACM Trans. Graph (2006)
9. Burt, P.J., Adelson, E.H.: The Laplacian Pyramid as a Compact Image Code. IEEE Trans. on Comm. (1983)

Pixel-wise Background Segmentation with Moving Camera

Neeraj Mishra[1], Manas Kamal Bhuyan[1], T. Malathi[1], Yuji Iwahori[2], and Robert J. Woodham[3]

[1] Department of Electronics and Electrical Engineering, IIT Guwahati, India
neeraj1928@gmail.com, {mkb,malathi}@iitg.ernet.in
[2] Department of Computer Science, Chubu University, Japan
iwahori@cs.chubu.ac.jp
[3] Department of Computer Science, University of British Columbia, Canada
woodham@cs.ubc.ca

Abstract. This paper proposes a novel approach for background extraction of a scene captured by a moving camera. Proposed method uses a codebook, which is a compression technique used to store data from a long sequence of video frames. This technique has been used to construct a model which can segment out the foreground and that with using only few initial video frames as a training sequence. It is a dynamic model which keeps on learning from new video frames throughout its lifetime and simultaneously produces the output. It uses a pixel-wise approach, and the codebooks for each pixel are made independently. Special emphasis has been laid on the intensity of an image as the human eye is more sensitive to intensity variations. A two layer modelling has been performed where codebooks are passed from the cache to the background model after satisfying the frequency and negative run length conditions. Experimental results show the efficacy of the proposed method.

Keywords: Video Surveillance, Codebook, Background Modelling.

1 Introduction

Separation of foreground objects from a video sequence is the foremost task in video surveillance systems, tracking systems, etc. The tracking system gets boosted up significantly if we have a good reliable dynamic real-time background model. The most important aspect for a background model to be successful is that it should be able to handle a dynamic background. Additionally, a background model should efficiently handle a dynamic scene. The typical cases of moving tree leaves, water ripples, illumination change, etc must be handled by the model in a computationally efficient manner. Besides this, there may be chances of movement of the camera. In such cases, the model must possess a dynamic learning capability.

Earlier, unimodal background modelling was done for all the pixels [1]. Then, this approach was changed to mixture of Gaussians on all the pixels [2]. Models

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 423–429, 2013.

using Mahalanobis distance and image vectors were proposed in [3]. However, all these approaches fail for the scenarios involving fast variations in the background. Background modelling using codebook method was proposed in [4]. In this method, initially a long video sequence was used as a training data, which was stored in the compressed form at each pixel as codebooks. So, a static background model was formed. This model was able to handle the case of shadows and illumination change. But, it fails when any objects in the scene which was stationary during the training period moves afterwards. This happen because the model was totally dependent on the initial scenes. An improvement of it was given in [5], where after developing the model based on the training samples, another cache model was made. If any new object comes in front of the camera for a longer duration, then it will be included in the background. But a drawback here is that one has to wait till the training period to get over before getting the output, since the system is in the learning stage during that period of time. Also, it can not handle motion of the camera. A hybrid codebook model was proposed in [6] which uses the mixture of Gaussians. This model gives good results, however, for a setup involving a moving camera, it is unable to give the desired results. A model which considered the case of movable camera has been proposed in [7]. This model makes use of a homography matrix, however, it is computationally very complex.

To address some of these issues, a novel background modeling algorithm is proposed which can handle small motions of the camera with a lesser computational cost. The proposed technique does not require any dedicated training period as it is based on dynamic learning. The proposed algorithm is also seen to easily adapt to new scenes as the camera moves.

2 Proposed Method

In order to successfully deal with the situations wherein the camera shows some kind of movements, the proposed method handles each of the pixels separately. Initially the entire scene is taken as the foreground, and from that the model starts forming the codebook. Additionally, no dedicated time is allotted for the training period. The proposed method only makes use of 2-tuples $< \lambda_i, f_i >$. Also, unlike the existing algorithms [4], [5], Maximum Negative Run Length (MNRL) is not used in the proposed method. Instead, the proposed method makes use of only the latest negative run length λ of the codebook.

Initially the codebook is formed by considering two parameters $\lambda_i = 0$ and the frequency $f_i = 1$. This codebook is included in the background model only when the following conditions are satisfied:

$$\lambda_i < \epsilon_1, \qquad f_i > \epsilon_2 \tag{1}$$

The first condition in (1) indicates that for a codebook to be considered as the background, its negative run length must be less than some constant ϵ_1. The second condition implies that the object should remain in front of the camera for a certain minimum number of frames ϵ_2 to be considered as the background.

The proposed technique uses the YCbCr color space. As Y space accounts more for the luma component of the image, more tolerance has been given to it than the other two Cb and Cr channels which account for the chroma component of the image. This is done in order to counter the effects of sudden shadow or highlights due to illumination change. In the case of intensity change, the value of Y component mostly changes. Hence, although the object was supposed to be in the background comes in the foreground. Hence more tolerance has been given in the Y axis for learning the codebook than the Cb and Cr in YCbCr colour space.

Let $C = c_1, c_2, ..., c_i$ represent the codebook for the pixel consisting of L codewords. Each pixel has a unique codebook size based on its sample variations. Each codeword consist of YCbCr colour details and 2-tuples $aux_i =< \lambda_i, f_i >$. For any pixel x_t at time t, $x_t =< Y_t, Cb_t, Cr_t >$. The flowchart of the proposed algorithm is shown in Fig. 1.

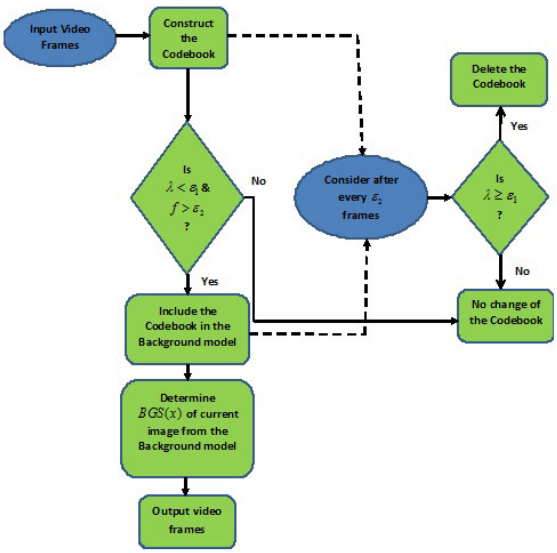


Fig. 1. Proposed Algorithm

Handling of Camera Motion: When the camera moves, the values at each pixel starts changing. As the pixel starts getting new values, it will be considered as the foreground. The background is defined as something which should be in front of a camera for at least ϵ_2 frames. Also at the same time, its negative run length should be less than ϵ_1 *i.e.,* the codebook should not be absent at that pixel for that number of frames. Generally at a particular pixel, we have 2 to 6 codebooks. For each frame, at each pixel, the λ and f value of all the codebooks are modified which are present at that pixel. First, the match with the value of the current pixel is checked, if the match is found then the λ is

assigned as zero and the f is incremented by 1. For the remaining codebooks, λ is increased by 1 and f remains the same. If there is no match, then a new codebook is formed. Additionally, the λ value of all the earlier codebooks will be incremented by 1, while the f value remains the same. This is the structure of the proposed cache model. Then in this model, at each pixel, the codebooks which satisfy the condition of the background are searched. Those which can satisfy this requirement are inducted in the background model B codebooks. Subsequently, the background subtraction of the current frame is implemented from this B model. Also after every ϵ_2 frames, the λ values of all the codebooks at each pixel in both the models are checked and those having values greater than ϵ_1 are deleted as there is no need of storing the values which are not repeating itself at the pixel. This method can efficiently handle very slow camera motions in any directions. The proposed method for the construction of codebook is detailed as follows:

Proposed Algorithm for Foreground Extraction

I $L \leftarrow 0$, $\boldsymbol{C} \leftarrow \phi$ (empty set)

II **for** $t \geq 1$ **do**

i. $x_t = (Y_t, Cb_t, Cr_t)$

ii. Find the codeword $\boldsymbol{c_m}$ in $\boldsymbol{C}$=$[\boldsymbol{c_i}|[1 \leq i \leq L]$ matching to x_t based on following two conditions:
(a) colordistance: $|Cb_t - Cb_m| \leq \delta_1$, $|Cr_t - Cr_m| \leq \delta_1$
(b) brightness: $|Y_t - Y_m| \leq \delta_2$

iii. If $\boldsymbol{C}$=ϕ or there is no match then $L \leftarrow L + 1$.
Create a new codeword $\boldsymbol{c}_L$ by setting:
$v_L \leftarrow (Y, Cb, Cr)$
$aux_L \leftarrow (0, 1)$

iv. Else update the matched codeword $\boldsymbol{c_m}$ consisting of
$v_m = (Y, Cb, Cr)$ and $aux_m = (\lambda_m, f_m)$ by setting:
$v_m \leftarrow \left(\frac{f_m \bar{Y}+Y}{f_m+1}, \frac{f_m \bar{C}b+Cb}{f_m+1}, \frac{f_m \bar{C}r+Cr}{f_m+1}\right)$
$aux_m \leftarrow \langle 0, f_m + 1\rangle$

III For each codeword $\boldsymbol{c_i}, i = 1, \ldots, L$
If $f_m > \epsilon_1$ & $\lambda < \epsilon_2$ then move that codeword in the Background Model **ß**

IV Background Subtraction

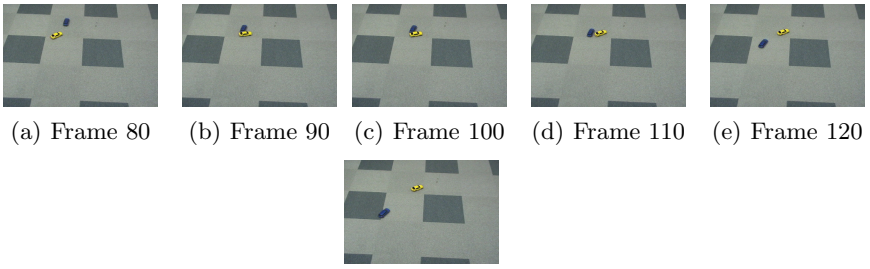
i. $\boldsymbol{x} = (Y, Cb, Cr)$

ii. For all codewords in Background Model **ß**, find the matching codewords based on two conditions:
(a) colordistance: $|Cb_t - Cb_m| \leq \delta_3$, $|Cr_t - Cr_m| \leq \delta_3$
(b) brightness: $|Y_t - Y_m| \leq \delta_4$
Update the matched codeword as in Step II iv.

iii. $BGS(x) = \begin{cases} \text{Background}, & \text{if match found} \\ \text{Foreground}, & \text{otherwise} \end{cases}$

V After every ϵ_2 frames, all codebooks having $\lambda \geq \epsilon_1$ will be deleted.

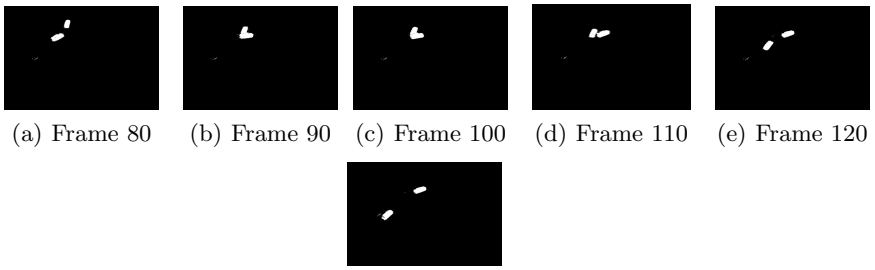
In this, the thresholds $\delta_1, \delta_2, \delta_3$ and δ_3 are judiciously selected to measure colour and brightness distortions for a particular video sequence.



(a) Frame 80 (b) Frame 90 (c) Frame 100 (d) Frame 110 (e) Frame 120

(f) Frame 130

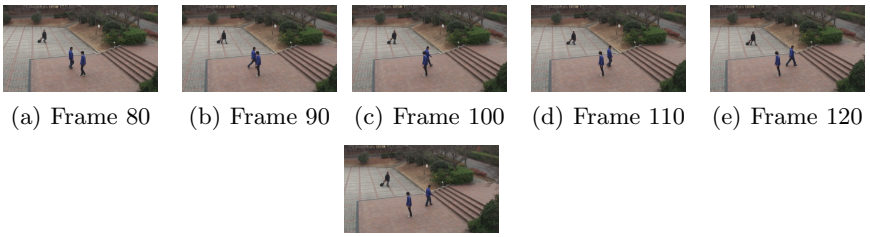
Fig. 2. Input video frames



(a) Frame 80 (b) Frame 90 (c) Frame 100 (d) Frame 110 (e) Frame 120

(f) Frame 130

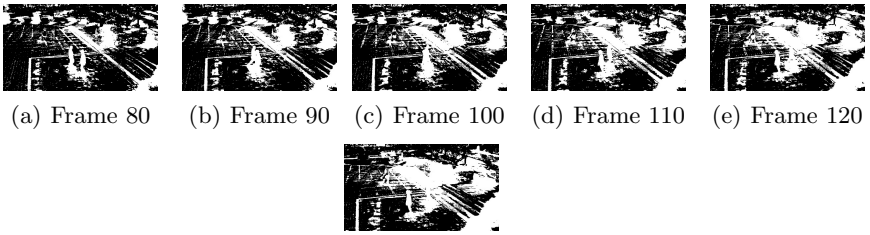
Fig. 3. Detection performance of the proposed method for input images in Fig. 2



(a) Frame 80 (b) Frame 90 (c) Frame 100 (d) Frame 110 (e) Frame 120

(f) Frame 130

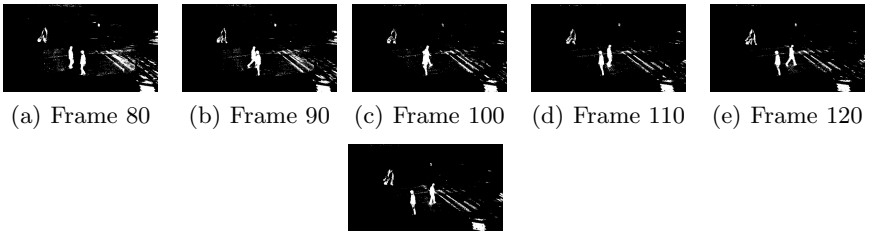
Fig. 4. Input video frames



(a) Frame 80 (b) Frame 90 (c) Frame 100 (d) Frame 110 (e) Frame 120

(f) Frame 130

Fig. 5. Detection performance of the existing method for input images in Fig. 4



(a) Frame 80 (b) Frame 90 (c) Frame 100 (d) Frame 110 (e) Frame 120

(f) Frame 130

Fig. 6. Detection performance of the proposed method for input images in Fig. 4

3 Experimental Results

The proposed algorithm has been made to run on the Core 2 Duo E8400 3.00GHz processor with 4.00 GB RAM. The size of the video frames is 720 x 480 pixels. The proposed method has been tested for two different scenarios involving a static and a moving camera. The results obtained have been detailed under Case Study 1 and Case Study 2 respectively. With static camera (Case 1), the speed of proposed algorithm is around 37ms/frame while with the moving camera (Case 2) it is around 47ms/frame. Values of ϵ_1 and ϵ_2 are selected as 25 and 50 respectively.

Case Study 1 (Static Camera): In this, Fig. 2 shows the input images, and Fig. 3 demonstrates the foreground detection performance using the proposed algorithm. It is observed that the proposed approach does not have unwanted small noises, unlike the results from earlier codebook methods in [4], [5]. Usually, the earlier methods make use of post-processing techniques like the 8-connected component labeling on the output in order to remove these unwanted small noises.

Case Study 2 (Moving Camera): For this, a moving camera is taken to capture the videos and the camera moves very slowly in any directions. The input images of the video sequence are shown in Fig. 4. The detection performance using the existing methods [4], [5] is given in Fig. 5. Finally, Fig. 6 demonstrates the detection performance of the proposed algorithm. The proposed method clearly outperforms the existing methods.

4 Conclusions

The proposed foreground background segmentation method is able to detect the foreground moving objects without making use of long sequences of video frames as a training data. Unlike previous methods, the parameters being passed to the codebook element are only the frequency f, and the negative run length λ which has resulted in a significant reduction in the computation time and better memory utilization, thereby enabling effective real-time implementation. Also, by giving extra allowance to the luminance component than chrominance,

the shadow effects has been countered efficiently. The efficacy of the proposed method for real-time background segmentation in the scenarios involving a moving camera has also been established.

Acknowledgement. Iwahori's research is supported by JSPS Grant-in-Aid for Scientific Research (C) (23500228) and Chubu University Grant. Woodham's research is supported by the Natural Sciences and Engineering Research Council (NSERC).

References

1. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real- time tracking of the human body. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 780–785 (1997)
2. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 246–252 (1999)
3. Seki, M., Fujiwara, H., Sumi, K.: A robust background subtraction method for changing background. In: 5th IEEE Workshop on Applications of Computer Vision, pp. 207–213 (2000)
4. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Background modeling and subtraction by codebook construction. In: Proc. of the IEEE Conf. on Image Processing, pp. 3061–3064 (2004)
5. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. Real-time Imaging 11(3), 172–185 (2005)
6. Sun, I.-T., Hsu, S.-C., Huang, C.-L.: A hybrid codebook background model for background subtraction. In: IEEE Workshop on Signal Processing Systems (SiPS), pp. 96–101 (2011)
7. Jin, Y., Tao, L., Di, H., Rao, N.I., Xu, G.: Background modeling from a free-moving camera by multi-layer homography algorithm. In: Proc. of the IEEE Conf. on Image Processing, pp. 1572–1575 (2008)

A Fast Video Inpainting Technique

Mrinmoy Ghorai, Pulak Purkait, and Bhabatosh Chanda

Indian Statistical Institute
Kolkata, India
{mgre04,pulak.isi}@gmail.com,chanda@isical.ac.in

Abstract. In this paper we present a fast video inpainting technique to infer the unknown information in the target region by maximizing box-based self-similarity and coherence measure. The video inpainting is already proposed in the literature and some of them are able to produce good quality results. However, the bottleneck of those algorithms is they are painfully slow. Here we fill the texture in the target region that preserves the smooth motion of the object without inclusion of any artifacts in reasonable amount of time. Our experiments show that the proposed method is quite efficient to synthesize unknown information in a video and comparable to the existing state-of-the-art methods. Moreover, proposed method is based on box filling and optimization is done on multiple scale using EM algorithm, and is computationally faster than the existing ones.

Keywords: video inpainting, self-similarity, spatio-temporal coherency.

1 Introduction

Video inpainting or completion is a method to fill-in the unknown regions in a video or image sequence. This is an extension to image inpainting problem in some sense where fill-in occurred in an unknown region in the place of the object we wish to remove from an input image. But here in video inpainting, target region placed in all the frames of the video at same location or continuous moving location. We can think the target region as a "hole" or "tunnel" in the video. In video inpainting techniques, the hole is filled by the known matter in such a way that it possess a smooth transition of the object having a motion and passing through the hole. Here we develop similar kind of method people use for image inpainting.

Approaches to image inpainting may be divided roughly into three categories: (i) partial differential equation (PDE) based approach for structure propagation, (ii) exemplar-based approach for texture synthesis and (iii) coherency-based approach for global consistency. PDE-based approaches [1] fill the target region of an image by diffusing the known data from the source region towards the interior of the target region. The exemplar-based approach infer unknown information by copying most similar patch from the source region and filling in the unknown region [2,3,4]. In the process of copying patches to the target region, a number of authors suggested to incorporate some spatial coherence in the texture synthesis process [5,6]. The basic idea of this type of approach is to assign each pixel in the target region based on the correspondence of the neighbouring pixels. Bugeau *et al.* [7] proposed a combination of three previously

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 430–436, 2013.

mentioned methods, namely texture synthesis, diffusion (PDE) and coherence in a single framework to take the advantages of all the methods. In this paper, we focus on exemplar based video inpainting method that extend exemplar-based image inpainting methods.

Wexler *et al.* [8] are first to explore region completion for video. They solved the video completion problem as optimization of global objective function using coherence structure. They used an iterative method for filling-in each pixel of the target region in multiple scale. Similar kind of work for repairing damaged video has been reported in [9]. Though this method is able to produce good results in difficult cases, their method involve combining different techniques making the process of inpainting very slow and complicated. The algorithm in [10,11] estimates the motion information for each pixel in the video frame in order to determine whether a pixel belongs to a moving object or belongs to static background accordingly inpaint the moving objects based on the information at stationary background. Their assumption was that the target region is much smaller than moving object size, having stationary background (static camera) which may not be the case in real scenario. This method is comparatively fast but not up-to that level.

In this paper, we extend the method of self similarity and coherency proposed in various image inpainting techniques to video inpainting and try to develop an algorithm which is computationally efficient compare to the existing methods. Here we suggest to

- model the problem of video inpainting as a energy minimization task;
- incorporate the strength of self-similarity along with coherency in a comprehensive framework;
- Reduce the computational cost by incorporating 3D-box based processing and filling instead of one pixel at a time.

Our algorithm try to approximate to a global optimization problem that combines two fundamental concepts of self-similarity and coherency.

The rest of the paper is organized as follows. In the following Section 2 describes proposed 3D box-filling based coherent texture synthesis procedures in detail and more implementation issues are discussed in Section 5. At the end, we show some experimental results in Section 4 where some practical evidences are also presented and then conclude with mentioning some pros and cons in Section 5.

2 Proposed Video Inpainting

Here we want to infer the unknown/target regions in a given video from the known/source region without introducing any artifacts. The challenge of this type of problem is to complete target regions of the video sequence in such a way that it allows moving objects to transmit smoothly through the target region. If the target region is smaller compared to the moving object then difficulty is much less than when the target region is quite larger than the moving object. Traditional image inpainting technique in general fails to generate temporal information of the moving object.

Let V and $T \subset V$ be a video and a target region either fixed or continuously moving throughout the sequence. Since video is a space-time volume, we define spatio-temporal

3D box $S_p(x,y,t)$ of size $(n \times n \times n)$ centering at the location of a pixel (x,y) in the t^{th} frame. Let us denote the set of all 3D box (full or partial) in the target region T by Δ and set of all box in the whole sequence V by Γ. We extend the texture synthesis technique proposed by [7] for video by finding the *correspondence map* $\tau : \Gamma \to \Gamma \setminus \Delta$ that associates the boxes from V to the boxes of known portions $V \setminus T$ such that

$$\tau(S_p) = \begin{cases} S_p, & \text{if } S_p \in \Gamma \setminus \Delta \\ S_q \in \Gamma \setminus \Delta, & \text{if } S_p \in \Delta \end{cases} \tag{1}$$

where S_q is a 3D box from known region, most similar to the box S_p of target region. If we can able to find out an efficient correspondence map τ that can map each box of the target region to known region, then we can fill the boxes in the target region by the mapped boxes in the source region. Thus we eventually solve the inpainting problem. Therefore in the proposed method we seek for the optimal relevant corresponding map.

2.1 Coherency and Self-similarity

In iterative texture synthesis or image inpainting, self-similarity is used to refine the assignment of patches or, more accurately, pixels in each iteration. The idea is to use the output at the previous iteration as input for the current iteration. Efros *et al.* [12] proposed texture synthesis for images in the terms of correspondence map τ. Note that here we use τ as a correspondence map of boxes $S_p \in \Delta$. In the proposed method, we assume that our target portions T of the given video V has self-similarity and coherency with known parts of the video. In other words, we wish to complete the target portions T with some new data $\widehat{T}$ such that resulting video $\widehat{V}$ has as much self-similarity and global visual coherence as the source portions $V \setminus T$. Therefore, we seek a solution of the following maximization problem

$$\widehat{\tau} = \max_{\tau} [Coherence(\widehat{V}\ ,\ V \setminus T)] \tag{2}$$

where τ is the correspondence map. The $Coherence(\widehat{V},\ V \setminus T)$ is the measure of self-similarity and global visual coherence, defined as

$$Coherence(\widehat{V},\ V \setminus T) = \sum_{S_p \in \Delta} s(S_p, \tau(S_p)) \tag{3}$$

where $s(S_p, \tau(S_p))$ is the similarity measure between box around the voxel p and corresponding mapped box around voxel q (where $S_q = \tau(S_p)$). A formalism similar to eq. (3) was already used in [8] for summarizing visual data. In our case we have considered the similarity measure as:

$$s(S_p, \tau(S_p)) = \exp\left(\frac{-d_{SSD}(S_p,\ \tau(S_p))}{2\sigma^2}\right), \tag{4}$$

where $d_{SSD}(S_p,\ \tau(S_p))$ computes the sum of square differences (SSD) of the pixel values between boxes S_p and $\tau(S_p)$. In the proposed method, we have considered some

features to represent the boxes instead of simple pixel values (RGB). The detail description of the features are mentioned in Section 3.2. The parameter σ is chosen manually to a high value. However, the choice is not critical for producing good output.

As we discussed, solution to the video inpainting problem would be given by the corresponding map τ that maximizes (2) defining self-similarity and coherency. Since the optimization problem (2) involves a non-linear objective function (NP hard), it is difficult to solve in a straight-forward method way. So we propose to proceed with iterative Expectation-Maximization (EM) algorithm. First, we initialize the correspondence map τ by random guess. In E-step, we generate target texture according to given correspondence map τ based on the source texture, and in the following M-step, we update the current guess of τ by assigning the boxes in the target region to the boxes in the source region. The E-step and M-step inherently maximizes the coherency and self-similarity of the filled region with the known portion respectively. We enforce self-similarity in the correspondence map τ by assigning nearby boxes by their neighbouring boxes. The coherence between boxes in T and those in rest of the video $V \setminus T$ as shown in eq. (2) is maximized if for every box $S_p \in \Delta$ all the surrounding boxes $[S_{p_1}, S_{p_2}, \ldots, S_{p_k}]$ agree on the box assignment at $\tau(S_p))$ with all the corresponding location of $[\tau(S_{p_1}), \tau(S_{p_2}), \ldots, \tau(S_{p_k})]$ appear in the video $V \setminus T$. Therefore, the iterative E-step aims to satisfy this condition for every box $S_p \in \Delta$, and the M-step searches for the best similar box in the box subspace $\Gamma \setminus \Delta$ of the unaltered region $V \setminus T$ of the video V. Let $[S_{q_1}, S_{q_2}, \ldots, S_{q_k}]$ denotes the boxes in $V \setminus T$ that are most similar to $[S_{p_1}, S_{p_2}, \ldots, S_{p_k}]$. Then the predicted S_{p_i} would be reliable if $s_i = s(S_{p_i}, S_{q_i}) \approx 1$. Therefore ,at each iteration, for each box $S_p \in \Delta$ and corresponding surrounding box S_{p_i}, we need to find out best possible box S_{q_i} in $V \setminus T$. Then we replace the box at S_p by the weighted average of the box at the corresponding locations of the similar boxes S_{q_i}. The weights are simply taken as the similarity measure s_i between the corresponding boxes S_{p_i} and S_{q_i}. Now the required huge computation for searching process to find the nearest neighbour (most similar) is reduced significantly by compensating it to approximate nearest neighbourhood [13]. This procedure may be expressed explicitly in the following way

E-step:

$$\widehat{V} := V \setminus T \cup \widehat{T} \tag{5}$$

where $\widehat{T}$ is obtained from T by replacing each 3D-box S_p by $\tau(S_p)$. Usually the boxes are overlapping and the boxes are aggregated in overlapping region. The aggregation is done by the weighted average of the overlapping boxes where the weights come from the similarity measure (SSD) between boxes S_p and $\tau(S_p)$.

M-step:

$$\begin{aligned} \widehat{\tau} &:= \arg\max_{\tau} \textstyle\sum_{S_p \in \Delta} s(S_p, \tau(S_p)) \\ &:= \arg\min_{\tau} \exp\left(\frac{-d_{SSD}(S_p,\ \tau(S_p))}{\sigma^2} \right) \end{aligned} \tag{6}$$

where $\widehat{\tau}$ is the modified estimation of τ. At each iteration we update the target region T and for each box $S_p \in \Delta$ (full or partially unknown box space), we find most similar boxes (candidates) $\tau(S_p)$ as the box $S_q \in \Gamma \setminus \Delta$. If there are several boxes minimizing this quantity we select one arbitrarily. The iterative process should end when the correspondence map $\widehat{\tau}$ remains unchanged in two consecutive iterations, i.e., $\tau(S_p)$ assigns same $S_q \in \Gamma \setminus \Delta$, $\forall S_p \in \Delta$ in two consecutive iterations. However, the solution of the

maximization problem may not converge to the actual global minimum or to a stationary point. As there is no guarantee that the iterative process converge to a stationary point, we set the stopping criterion as the maximum number of iterations.

3 Implementation Details

The algorithm is applied on a video V with target region T where inside T information is missing and the algorithm is supposed to infer missing information from the rest of the video $V \setminus T$. The proposed algorithm is an iterative method and in each iteration 3D boxes within T are refined according to the maximization of coherency and self-similarity.

Nearest neighbour search is an important problem in a variety of applications, including the knowledge discovery and data mining, pattern recognition and classification. The important task is to build a fast algorithm that could able to find nearest neighbour of every patches of an image. Obviously the problem can be solved in $O(dn)$ time through simple brute-force search for n points in d dimensional space. We use an efficient optimal algorithm for approximate nearest neighbour search [14] that can solve the problem in $O(d \log n)$ or less time.

Instead of using discrete patches, we use overlapping ones and use simple weighted average over the overlapping portions where the weights are computed as proportional to the similarity measure of the patches.

3.1 Multiscale Implementation

To enforce the global consistency further and also to speed up convergence, we perform the iterative process in multiple scales in both spatial and temporal directions. Each of the scale makes the resolution a fraction of the resolution of the upper scale. Scaling factor $1.25 - 2.00$ can produce significant result in most of the cases. In all of our experiments, we have chosen resolution scale to be 1.5 in both the directions. The optimization is done by using EM technique starting at coarsest scale and the solution is propagated to finer levels for further refinement. Initially, we fill the the target region for each frame of the video by some random texture at the coarsest scale followed by a few EM iterations. The filled region gets refined as iteration goes. Then, both T and $V \setminus T$ are gradually upsampled to finer resolutions, followed by more EM iterations, until the final fine resolution is obtained. We fix the number of iteration as the stopping criterion to terminate the process. However, we can terminate during the iterative process at the topmost scale when we achieve acceptable visual quality.

3.2 Spatio-temporal Similarity Measure

The Sum of Square Differences (SSD) of color channels is widely used in image inpainting but it is inadequate to produce the desired result in spatio-temporal space. Since human visual system is very sensitive to motion, a well-behaved measure needs to acknowledge human visual *perception*. For this we would like to incorporate some motion information into our algorithm. We add optical flow in our measure to obtain

motion information. Suppose a pixel at location $p = (x,y,t)$ in one frame moved to $(x+\partial x, y+\partial y, t+\partial t)$ in the next frame. Then $v_x(p) = \partial x/\partial t$ and $v_y(p) = \partial y/\partial t$ gives the motion estimation at p. If the motion is only in the horizontal direction, then u_x captures the instantaneous motion in the x direction. Similarly for vertical direction, v_y captures instantaneous motion in the y direction. These two measures depend upon the spatial and temporal changes while capturing object velocities. We add these two components after scaling the RGB values to obtain a five-dimensional representation for each space-time point: (R,G,B,v_x,v_y). We apply SSD to this $5D$ feature vector to capture spatial and temporal similarities simultaneously. So for two space-time boxes S_p and S_q, we have $d_{SSD}(S_p,S_q) = \|u(S_p) - u(S_q)\|_2^2$. We take nearest neighbours of each box in the target portion T using the distance measure d_{SSD} on the 5D representation of space-time points and keep it to update the pixels in the target region.

Fig. 1. Results of proposed video inpainting method. The first column contains some frames of the input video where the umbrella is removed during video completion. The second column are the result of Wexler *et al.* [15] and the right column represents resultant frames using proposed method.

4 Experiment and Results

In this section experimentally we set different parameters used in our algorithm and test it for some videos. In our experiment, the size of the box is chosen as $5 \times 5 \times 5$ with 3-pixel overlap. For faster synthesizing process, the number of scale in the pyramid is chosen as 4 with resolution factor 1.6 and the number of iterations in each scale is chosen as 10.

In Fig. 1, we displayed the result of proposed video completion techniques. We have displayed only some frames to demonstrate the motion preservation of the object while filling the target hole. We observe comparable result with Wexler *et al.* [15] and no additional artificial distortions occurred during box-based filling. This result is generated within 20 minutes where as Wexler *et al.* had reported more than 10 hrs to get that output. with other videos also we experience these faster completion.

5 Conclusion

In this paper we describe a fast video inpainting algorithm by maximizing box-based self-similarity and coherency in a comprehensive framework. We combine these two concepts into a maximization problem and optimize by EM-algorithm to produce the inpainted video. Experimental results show that the output of the proposed method is comparable to existing approaches.

References

1. Bertalmio, M., Sapiro, G.: Image In painting. In: Proc. of the ACM SIGGRAPH Conf. on Computer Graphics, New York, USA, pp. 417–424 (2000)
2. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, pp. 341–346. ACM, New York (2001)
3. Criminisi, A., Perez, P., Toyama, K.: Object Removal by Exemplar-Based Image Inpainting. In: Proc. IEEE Int. Conf. on Computer Vision, vol. 2, pp. 721–728 (2003)
4. Sun, J., Yuan, L., Jia, J., Shum, H.: Image Completion with Structure Propagation. In: Proc. SIGGRAPH 2005, pp. 861–868 (2005)
5. Ashikhmin, M.: Synthesizing natural textures. In: Proc. of ACM Symp. on Interactive 3D Graphics, pp. 217–226 (2001)
6. Komodakis, N., Tziritas, G.: Image Completion using global optimization. In: Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition, pp. 442–452 (2006)
7. Bugeau, A., Bertalmio, M., Caselles, V.: A comprehensive framework for image inpainting. IEEE Transaction on Image Processing 19 (October 2010)
8. Wexler, Y., Shechtman, E., Irani, M.: Space-time video completion. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. I.120–I.127 (2004)
9. Jia, J., Pang Wu, T., Wing Tai, Y., Keung Tang, C.: Video repairing: Inference of foreground and background under severe occlusion. In: Proc. Computer Vision and Pattern Recognition, pp. 364–371 (2004)
10. Sapiro, G., Patwardhan, K.A., Bertalmio, M.: Video inpainting of occluding and occluded objects. In: IEEE Int. Conf. Image Processing., vol. 2, pp. 69–72 (2005)
11. Patwardhan, K.A., Sapiro, G., Bertalmio, M.: Video inpainting under constrained camera motion. IEEE Transc. on Image Processing 16, 545–553 (2007)
12. Efros, A.A., Leung, T.: Texture synthesis by non-parametric sampling. In: IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 1033–1038 (1999)
13. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics, SIGGRAPH, 24.1–24.11 (2009)
14. Arya, S., Mount, D.M., Netanyahu, N.S., Silverman, R., Wu, A.Y.: An optimal algorithm for approximate nearest neighbor searching fixed dimensions. Journal of ACM 45, 891–923 (1998)
15. Wexler, Y., Shechtman, E., Irani, M.: Space-time completion of video. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 463–476 (2007)

Static Summarization of Video Scenes Based on Minimal Spanning Tree

Partha Pratim Mohanta[1], Sudipta Chowdhury[2], Arnab Roy[2], Sanjoy Kumar Saha[2], and Bhabatosh Chanda[1]

[1] ECS Unit, Indian Statistical Institue, Kolkata, India
[2] CSE Dept., Jadavpur University, Kolkata, India

Abstract. The common practice for providing a static summarized view of a video is to create a storyboard. Storyboard is the chronological arrangement of the representative frames. Shot level storyboard suffers from redundancy as in a scene constituting shots normally repeat. Also the size of such storyboard is a constraint for many application. In this work we have considered scene as the more meaningful unit. We propose a state-based scene segmentation algorithm and also a minimal spanning tree based novel method to select the representative frames for the scenes. Storyboard consisting of scene level representative frames are much more compact than shot level storyboard. Moreover, scene being the semantic unit, flow of semantic content of the video data is well preserved. Experimental result confirms the claim.

1 Introduction

The huge volume is the major challenge in working with video data. As a result efficient browsing, indexing and summarization of video data has emerged as an active area of research in the recent years. For all such application, video segmentation acts as a fundamental step. Video segmentation is the process to partition the video into meaningful units. Shot is the basic structural unit and it is a sequence of frames captured through single operation of a camera. But for better understanding a video, a higher level semantic unit, called *scene* is formed. A scene is a collection of semantically related adjacent shots which depicts a high level concept.

The two major tasks for storyboard based summarization are video segmentation and thereafter determining the representative frames for the segmented units. In the context of video segmentation most of the efforts are directed towards temporal segmentation *i.e.* shot detection. Transition between the shots may be abrupt (cut) or gradual. A lot of works have been reported for shot level segmentation [1,2]. Past study reveals that so far the summarized view of a video has been considered as a storyboard/ToC at the shot level. Semantically scene is the unit of the story conveyed by a video data. Ideally, the scene level storyboard can represent the content much better rather than shot level storyboard or a reduced version of it. A scene consist of contiguous shots that are semantically

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 437–444, 2013.

related, but may not be visually similar. Shots in a scene usually follow a quasi-periodic repetition. A heuristic approach based on the quasi-periodic pattern of the shots [3] is presented to detect the scenes in a video. Various Graph based schemes are also reported in [4,5].

The next crucial task is to determine the representative frames (or key-frames) of shots. Determining the number of key-frames in a shot is an issue. Early approaches [6,7] considered first/last frame of the shot or both as representative. Mohanta et al. [8] have proposed a scheme that decomposes a shot into sub-shots with uniform visual content and from each such sub-shot one frame is taken as the representative. Using these key-frames a shot level storyboard may be generated in [9]. As the similar shots repeat in a scene, shot level storyboard bears substantial redundancy. Fuzzy-ART and Fuzzy C-Means based scheme [10] are also tried to minimize the redundancy. Mohanta et al. [8] have presented a spanning tree based scheme to prepare a size constrained storyboard.

Video summarization has many important applications including video-on-demand. Summarized video in the form of storyboard consisting of representative frames are most appealing and convincing compared to any textual description. Such summarization is most efficient and effective if it is done at the scene level. in this work we summarize a given scene by a minimal subset of representative frames. These frames placed in chronological order forms the storyboard and also act as bookmark/table of content (ToC) enabling random access to the desired unit. Our main contribution in this work are two folds: first, we have modified algorithm presented in [3] for scene detection; second, is to map the classical MST algorithm in a novel way to divide the key-frames of a scene into visually homogeneous groups and also to extract representative frame from each of these groups. The present endeavor is a continuation of our earlier work on shot segmentation [2], key-frame extraction [8] and scene segmentation [3]. However, care is taken to design the proposed algorithm such that any error committed in the earlier stages does not affect the output much.

The paper is organized as follows. Following this introduction and a brief survey of the relevant work, a scene detection algorithm which is a modified version of our earlier work [3] is presented in section 2. Then, in the same section we present a methodology to find out the scene level representative frames which will be used in the storyboard. Experimental results are presented in section 3 and finally, concluding remarks are cited in section 4.

2 Proposed Methodology

Scene based summarization is more meaningful where shots in a scene together represent an event in the context of the semantic content of the video. Although the shots may not be visually uniform but they jointly contribute to narrate the story in the scene. It has a strong resemblance with the human perception [11]. Representative frames of the scenes may be organized to provide a summarized view/ToC at the scene level. Moreover, such a view is free from the redundancy present in the shot level storyboard and renders a more concise representation.

The output of earlier schemes which work on shot level storyboard do not ensure the continuity in terms of the story depicted by the video. But, scene level storyboard ensures continuity of the events and also achieves conciseness. This dual advantage has motivated us to go for scene level summarization. The major steps of scene level summarization are as follows.

- Segment the video data
  - Detect the shot boundaries
  - Identify the representative frame(s) for each shot
  - Detect the scene boundaries
- Identify the representative frames for each scene
- Organize the representative frames into storyboard

2.1 Shot and Key-Frame Detection

In our work, shot boundaries are detected following the methodology presented in [2]. A unified model of shot boundary detection has been presented to deal with both abrupt and gradual transition. To detect the scene boundaries, each shot is first represented by one or more representative frames (key-frames). For each shot, key-frame(s) are selected using the methodology presented in [8]. In this work, Wald-Wolfowitz run (WWR) test is applied on each shot to examine its consistency in terms of visual content. The hypotheses of the WWR test are stated as follows:

H_0: X and Y are from same population, *i.e.,* $F_x = F_y$

H_1: They are from different population, *i.e.,* $F_x \neq F_y$

where X and Y are two samples of sizes m and n respectively and the corresponding distributions are F_x and F_y. Test statistic W is computed based on R, the number of runs (a '*run*' is a sequence of identical labels) as follows.

$$W = \frac{R - \frac{2mn}{N} - 1}{\sqrt{\frac{2mn(2mn-N)}{N^2(N-1)}}} \tag{1}$$

where $N = n + m$. Each frame is represented by a feature vector. Thus the multivariate version of the WWR test is used to detect the shots (sub-shots) which are visually consistent. For each such shot (sub-shot) one frame is taken as the representative based on the fidelity criteria.

2.2 Scene Detection

Shots in the video are represented by their key-frames and the video data becomes a sequence of key-frames. Scene boundaries are then detected by a modified technique of what is presented in [3]. For this we adopt the bag of words model [12]. Key-frames are clustered and replaced by the corresponding cluster tag. Thus, the video data is converted into a sequence of tags. This sequence is then analyzed to detect the scene boundaries. Shots in a scene repeat themselves with different periodicity. Thus, a scene is a quasi-periodic sequence of

shots. Based on this observation, concept of stable and quasi-stable state around which scenes are formed are defined. Finally, considering these states a heuristic algorithm is used to localize the scene boundaries as follows.

- **Input:** Let T_i $(i = 1, 2, \ldots, N)$ are the tags (or key-frames) representing the video, where $T_i \in \{t_i | i = 1, 2, \ldots, k\}$, k is the number of cluster of key-frames and $k << N$. Let p be the length of the quasi-period at the L-th iteration.
- **Step 1a:** Any tag T_i is marked Repetitive if there exists a tag T_j within the span $i \pm p$ such that $T_i = T_j$.
- **Step 1b:** If the number of Repetitive tags is less than $m_g \times N$ $(m_g < 1)$, increase p and repeat step 1a.
- **Step 2:** Suppose there exist two Repetitive tags T_i and T_j such that number of repetitive tags from T_i to T_j inclusive is more than $m_l \times |j - i|$ $(m_l < 1)$, we call this span of tags as *stable state*.
  If a run of tags, which is not marked as stable state, is bounded by two stable states $\hat{S}_1$ and $\hat{S}_2$ and run length lies between p and $2p$, we call this span of tags as *Quasi-stable state*.
- **Step 3a:** If a quasi-stable state Q is bounded by $\hat{S}_1$ and $\hat{S}_2$, and $\hat{S}_1$ and $\hat{S}_2$ possess some common tag, the $\hat{S}_1$, Q and $\hat{S}_2$ are merged to a scene.
- **Step 3b:** If 3a is not true, and Q and $\hat{S}_1$ possess some common tag then $\hat{S}_1$ and Q merged to a scene.
- **Step 3c:** If 3a is not true, and Q and $\hat{S}_2$ possess some common tag then Q and $\hat{S}_2$ merged to a scene.
- **Step 3d:** If 3a, 3b or 3c are not true, Q is divided into two halves and each half is merged with adjacent $\hat{S}$ to form a scene.
- **Step 4:** If no more merging possible, feed each run of tags separately to step 1 to 3 as input with modified N and p.
- **Step 5:** If $p = N$ and still some unmerged tags remain, declare every run of tags a scene.

2.3 Selection of Representative Frames

Once the scenes are detected, the representative frames for each scene have to be determined. Suppose, C_i be the i-th scene of the video and $\tilde{F}_i$ be the set of all key-frames corresponding to the shots in the scene C_i. Thus, the task is to find out a $rf_i \subset \tilde{F}_i$ which can well represent C_i and redundancy present in $\tilde{F}_i$ is also minimized substantially. In the proposed methodology we present a minimal spanning tree based scheme to which $\tilde{F}_i$ is provided as the input and rf_i is obtained as the output. Each frame $f_j \in \tilde{F}_i$ is represented by n-dimensional feature vector $V_j =< v_1, v_2, \ldots, v_n >$ and treated as a node in the minimal spanning tree. To start with a node is taken as a root. Gradually, all the nodes are inserted into the tree. In each iteration from the remaining nodes, one nearest to the tree is included as the child of the nearest node in the tree. Distance between the child and parent is taken as the weight of the edge/link between them. The steps for preparing the tree are as follows.

- *remaining_set*= set of all nodes(feature vector)

- Choose randomly a node (feature vector), nd_i as the root
- *included_set* = nd_i
- *remaining_set*= *remaining_set* - *included_set*
  - Repeat the following steps until *remaining_set* is empty
  - For each node, $nd_r \in$ *remaining_set*
    - distance with the tree, $d_k = min\{dist(nd_r, nd_{in})\}$ where, $nd_{in} \in$ *included_set*
  - Let $d_t = min\{d_k\}$ for the nodes $nd_{tr} \in$ *remaining_set* and $nd_{ti} \in$ *included_set*
  - Put nd_{tr} in the tree as child of nd_{ti}
  - Set weight of the link between nd_{tr} and nd_{ti} as d_t
  - Put nd_{tr} in *included_set*
  - Remove nd_{tr} from *remaining_set*

In our experiment, each frame (node) in $\tilde{F}_i$ is represented by 512 dimensional color histogram. Each of the R, G, B plane of the frame is divided into 8 bins and normalized histogram is formed. Similarity between two nodes (*i.e.* feature vectors) is computed following Bhattacharyya distance. Distance between two nodes, $dist(nd_p, nd_q)$ is taken as $1 - bhattacharyya\_distance(V_p, V_q)$ where V_p and V_q are the normalized color histograms of the frames corresponding to nd_p and nd_q respectively.

To find out the representatives of the set of frames $\tilde{F}_i$, our target is to split the set into disjoint subsets where the elements in a subset are visually similar. Then one element from each subset is chosen as the representative based on certain criteria. Once the frames in $\tilde{F}_i$ are organized into minimal spanning tree (mst_i), the selection process of the representative frames for the scene C_i proceeds as follows.

- $rf_i = \{\}$
- For each edge/link e_{pq} in the minimal spanning tree, mst_i
  - if (weight of $e_{pq} > th$) then drop the link /edge
- for each subtree st_j in mst_i
  - find the node nd_m with maximum children
  - $rf_i = rf_i \cup nd_m$
- rf_i is the representative frames for the scene C_i

The minimal spanning tree, mst_i is divided number of disjoint subtrees $\{st_j\}$. Weight of a link/edge stands for the dissimilarity between the frames corresponding to a parent node and child node in mst_i. If it exceeds a threshold (th) then edge is removed. By removing such edges, mst_i is splitted into number of disjoint subtrees where st_j consists of the nodes corresponding to visually similar frames in $\tilde{F}_i$. Finally, one representative from each subtree is chosen. The node with maximum children represents the frame having closest visual similarity with the majority frames in the subset and taken as the representative. In order to determine the value of th, for each node (frame in $\tilde{F}_i$) distance with its nearest node (frame) is first determined. Let, for node nd_p such distance is d_{min_p}. Then th is taken as $\mu_i + \sigma_i$ where μ_i and σ_i stand for $avg\{d_{min_p}\}$ and $std\_dev\{d_{min_p}\}$

respectively. Similar frames also reflect variation among themselves. A low value for th will put them into different subtrees. On the other hand, a high value may merge dissimilar frames also. The proposed value for th is a balanced one. $\mu_i + \sigma_i$ restricts the subtree to intrude into a dissimilar one but allows the subtree grow so that similar frames are included despite of variation among them.

3 Experimental Result

In order to carry out the experiment, we have worked with four movie videos namely *Hirak Rajar Deshe* (HRD) (a Satyajit Ray film), *Mission Impossible-I* (MI), *Beautiful Mind* (BM) and part of *Catch Me If You Can* (CMIYC). The details of the movies are shown in Table 1.

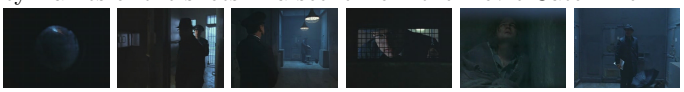
In our experiment shot and key-frames of a video are detected using the methodologies as described in section 2.1. Shot level storyboard is the arrangement of all the key-frames in order of their appearance in the video. Scene segmentation is performed by the algorithm presented in 2.2. Depending on the activity and dynamics present in the scene, it is a collection of different shots. Some of them are visually similar occurring at certain interval. Thus, shot level storyboard reflects lots of redundancy which are avoided in the scene level summary. Representative frames for the scenes are obtained following the methodology presented in the section 2.3. Chronological ordering of the same provides the summarized storyboard at scene level. Table 1 shows that in comparison to shot level storyboard, huge amount of size reduction is achieved in case of scene level summarization. Fig. 1 shows few examples where the key-frames of all the shots in a scene and representative frames of the scene are shown.

Table 1. Performance of summarization

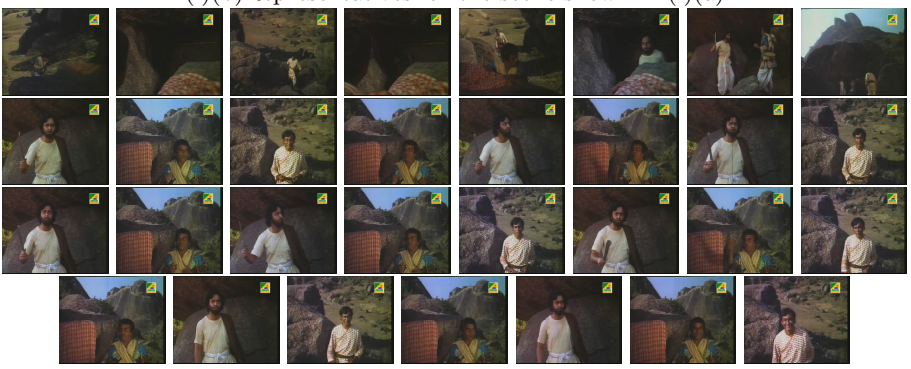
Movie	duration (HH:MM:SS)	# shots	# scenes	# frames in summary		% of reduction in scene level summary
				shot level	scene level	
HRD	01:52:27	1331	62	1511	113	92.52
MI	01:45:42	1465	121	1582	228	85.59
BM	01:55:39	1468	98	1768	172	89.93
CMIYC	00:36:37	312	32	427	48	88.76
Overall	06:10:25	4576	313	5288	561	89.39

Evaluation of summary is a difficult task as it is quite subjective. In order to do so, shot level storyboard of the video is presented to the viewers. Scene level summary generated by the system is then made available to them. Every viewer then marks which of the scene level representatives generated by the system are irrelevant and frames in the shot level storyboard which are desirable but missing in the scene level summary. With these information, precision and recall are computed. In our effort, we have considered 10 viewers. Average precision and recall achieved by the proposed methodology are 92.88% and 88.01% respectively.

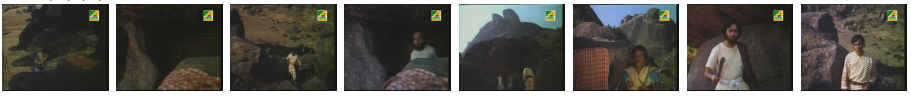
(i)(a) Key-frames of the shots in a scene from the movie Catch Me If You Can



(i)(b)Representatives for the scene shown in (i)(a)

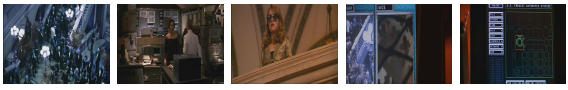


(ii)(a) Key-frames of the shots in a scene from the movie Hirak Rajar Deshe



(ii)(b)Representatives for the scene shown in (ii)(a)

(iii)(a) Key-frames of the shots in a scene from the movie Mission Impossible-I



(iii)(b)Representatives for the scene shown in (iii)(a)

Fig. 1. Few examples of scene level summary for scenes from different movies

4 Conclusion

In this work we have presented a state-based scene detection algorithm and a novel methodology to generate a scene level static summary of the video. In a shot level storyboard, key-frames of all the shots are presented chronologically. As similar shots appear periodically, the storyboard thus obtained possesses redundant frames. To get rid of it, we have considered scene as the semantic unit. A scene is a collection of consecutive shots of various types. A minimal spanning tree based scheme has been presented which determines the representative frames for the scene. Such representatives of all the scenes arranged chronologically renders the scene level static summary/storyboard. Experimental result indicates that the performance of the proposed methodology is satisfactory and achieves huge reduction in storyboard size in comparison to shot level summary.

References

1. Cernekova, Z., Pitas, I., Nikou, C.: Information theory-based shot cut/fade detection and video summarization. IEEE Trans. on CSVT 16(1), 82–91 (2006)
2. Mohanta, P.P., Saha, S.K., Chanda, B.: A model-based shot boundary detection technique using frame transition parameters. IEEE Transactions on Multimedia 14(1), 223–233 (2012)
3. Mohanta, P.P., Saha, S.K., Chanda, B.: A heuristic algorithm for video scene detection using shot cluster sequence analysis. In: Proc. ICVGIP, India (2010)
4. Rasheed, Z., Shah, M.: Detection and representation of scenes in video. IEEE Trans. on Multimedia 7(6), 1097–1105 (2005)
5. Zhao, Y.J., Wang, T.: Scene segmentation and categorization using n-cuts. In: Proc. CVPR, pp. 343–348 (2007)
6. Tonomura, Y., Akutsu, A., Otsugi, K., Sadakata, T.: Videomap and videospaceicon: Tools for automatizing video content. In: Proc. ACM INTERCHI, pp. 131–141 (1993)
7. Rui, Y., Huang, T.S., Mehrotra, S.: Exploring video structures beyond the shots. In: Proc. IEEE Conf. on Multimedia Computing and Systems, pp. 237–240 (1998)
8. Mohanta, P.P., Saha, S.K., Chanda, B.: A novel technique for size constrained video storyboard generation using statistical run test and spanning tree. International Journal of Image and Graphics 13(1), 1–24 (2013)
9. Macer, P.J., Thomas, P.J.: Video storyboards: Summarizing video sequences for indexing and searching of video databases. In: Proc. IEE Colloquium on Intelligent Image Databases, pp. 2/1–2/5 (1996)
10. Cayllahua-Cahuina, E.J.Y., Camara-Chavez, G., Menotti, D.: Video summarization with automatic shot detection using color histograms. In: Proc. IPCV 2012, Las Vegas, USA (2012)
11. Rui, Y., Huang, T.S., Mehrotra, S.: Constructing table-of-content for video. ACM Multimedia Systems 7(5), 359–368 (1999)
12. Sivic, J., Zisserman, A.: Efficient visual search of videos cast as text retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(4), 591–606 (2009)

Real-Time Smoke Detection in Video Sequences: Combined Approach

Malenichev Anton and Krasotkina Olga

Tula State University
Cybernetics faculty
Tula, Russia
malenichev@mail.ru, O.V.Krasotkina@yandex.ru

Abstract. This paper presents a combined approach for rapid smoke detection from video sequences using pre-improvement methods. Smoke is characterized by different properties like a color, irregularities in motion, smoothing the edges, etc. There are hard to describe it using basic image features. Usually smoke detection algorithms use some color and turbulence features for the smoke description. But our experiments shown the high false alarm rate of such algorithms. In our work we propose the additional smoke description features based on smoke transparency. Before the recognition we use the regions matching method to increase the flexibility of our system. As a first step in processing we extract background. Moving objects are candidates for smoke. The Gray World algorithm is used here. Compare the results with the original frames in order to get image features within some particular gray scale interval. After that we use the rate of color changing to checking of the transparency of current area. Last we calculate complexity of turbulent phenomena of the smoke shape and apply it to the incoming video stream. As a result we have just smoke regions on the video stream. There are different objects, shadows or illumination changes will not be mistaken for smoke by the algorithm. This method gives an early recognition of smoke in the observed scene.

1 Introduction

In many extreme situations there is visible smoke spread prior to flames. This fact can be of great help for high-efficiency fire control. Using video systems that are too sensitive to smoke can save many lives in dangerous situations. Time is a main factor for minimizing damages caused by fire and smoke. A short distance to the location of fire ignition is a crucial requirement for most of the traditional smoke sensors in order to be efficient. In many cases the fire is already in place when a chemical alarm is going off and the situation is out of control. Video smoke detection has low cost, high efficiency and low maintenance and becomes increasingly important because of it [1][4]. Video is a good option for large, open areas like forest fires, saw mills, petrochemical refineries, etc. Video is a volume sensor as opposed to a point sensor, which looks at just one point in space. Smoke or fire can be not at that point, so the smoke alarm would not be caused.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 445–450, 2013.

A volume video sensor has higher probability of early smoke or flame detection and monitors a larger area [7].
However, video smoke detections have some drawbacks. There are high rate of false alarm, the difficulty of defining smoke in terms of primitive image features such as motion, obscuration, intensity, edge, etc.
There are several smoke detection methods from video captured in visible-spectrum have been proposed. They use of color, motion and geometry of smoke regions. Then apply rule-based analysis or Bayesian analysis to detect the smoke. Out of the many approaches that address these issues some classification could be in place:

1. Using the fractal encoding concepts to find smoke areas from an image. This approach used the feature of smoke to self-similarity of shapes [5].
2. Definition of irregularities in motion due to non-rigidity of smoke. Here computed optical flow field using two neighbor images. Making use of entropy of the distribution of the motion directions It can be possible to separate smoked motion from non-smoked [8].
3. Extracting of image features such as flickering, motion, edge-blurring to segment moving, and edge-blurring regions out from video. The methods to extract these features were background subtraction, temporal wavelet transformation, and spatial wavelet transformation [10].
4. Extracting of local motions from cluster analysis of points in a multidimensional temporal embedding space in order to track local dynamic envelopes of pixels, and then used features of the velocity distribution histogram to discriminate between smoke and various natural phenomena such as clouds and wind-tossed trees that may cause such envelopes [11].
5. Using of different image features, such as the percentage of image change, correlation, variance etc., extracted from both background image and current image, and then used a rule-based or a rule-first-Bayesian-next analysis method to differentiate smoke motion from non-smoke motion [6].

As we can see, there are many approaches that address to these issues, but the main novelty of this work consists in the analysis the smoothing edges of objects beyond the smoke area.

First we make the RGB balancing of the each frame to obtain a real color level regardless of automatic white level adjustment, used in most video cameras. Next we use the color and motion threshold transformations by color and motion to determine regions of interest, which may contain the smoke. Next, most important step - checking the rate of change of color in areas of interest for N sequential frames. If a change in a particular pixel is more than a predetermined threshold, it is deemed opaque gray object. When the color change does not exceed the predetermined threshold, then we consider it a fact of the presence of smoke in the considered pixel. Next, if the percentage of pixels with smoke in a particular area of interest over a predetermined percentage, the area considered like a smoke. Finally we analyze the behavior of the area of the interest using the turbulence function and with high probability can determine presence or ab-

sence the smoke in current area. This simple and not time consuming algorithm allows with high accuracy detect smoke.

We aim to present a method that detects smoke in an early stage from video.

2 Detection of Smoke from Video

2.1 Areas Matching

During recognition the frame can contain multiple objects simultaneously. Therefore it is incorrect to apply the methods of dynamic analysis for all areas of interest as for a single area. We need to operate on local areas, not their totality. To solve this problem we use the areas matching method. Each region of interest in the frame may be regarded as:

1. First time appeared area (without parent area);
2. An area, changed its location and size (has one parent area);
3. An area, resulted as merging some parent areas;
4. An area, resulted as partitioning of the parent area;

The matching algorithm consist of following steps. Each of the M regions on the frame $(i+1)$ subjects operation of "logical and" with each of the N regions in the i frame and then areas of each intersection are determines. The result is stored in the matrix $R(m*n)$. The higher value of element correspond to the closer regions. Zero value means that areas does not intersect. Next we analyze the matrix R:

1. If all values in a row are zeros, then an area is new (without parent regions);
2. If a row have one non-zero value, then properties are inherited from one parent region;
3. If a row have more than one non-zero value, then the parent region is selected with the maximum area of intersection;

This algorithm has linear complexity, allow us to produce region matching easily and accurately and increase flexibility of our smoke recognition system.

2.2 Smoke Detection Algorithm

The first stage of video processing is a balancing of the color scheme of the frames. We calculate the average value of each R, G, B color component to receiving real level of gray color. It averages the RGB of the whole image and then it uses like a balancing constant.

The following stage of processing is a detection of a smoke regions on frames. This procedure is carried out in some steps. For smoke definition used chromatic characteristics of smoke areas. It is known that the smoke has color from light to dark gray. One of the ways to describe color feature of the smoke regions of an RGB image is in [3]:

$$|R-G| < T;\ |G-B| < T;\ |R-B| < T; \tag{1}$$

where T is a threshold, adjusted on training video set. Naturally, use of a chromatic characteristic for localization of smoke areas isn't enough. Moving of the smoke areas relatively to background frame can also be considered as the property of smoke. Therefore the following stage is detection of moving objects. Moving areas detection procedure is carried out by recalculating the background image in the sequence of frames. Next the moving objects calculated with subtraction of each color component of the image from background and dissectings away on a threshold. The threshold value is choosed by training sequence of videos. Potentially smoke areas are estimated as combination of moving areas and grey colored areas. Then we analyze the rate of change of color in them.

When we see a scene of smoke, the color of the objects behind it does not change instantly. As soon as the smoke condensed, the color of the object is gradually replaced by the color of the smoke. Let ΔR, ΔG, ΔB - the absolute change in the color of a pixel regions of interest for the N frames. Define the threshold of maximum color change C_{max}. The criterion of the presence of smoke in the area will be fulfillment of conditions:

$$\Delta R < C_{max};\ \Delta G < C_{max};\ \Delta B < C_{max}; \tag{2}$$

As we are dealing with smoke, its natural phenomena is that it strives to move up. This feature can be detected by monitoring positive local extrema of the turbulence (3).

$$\Omega = \frac{P(t)}{2\sqrt{\pi \cdot A(t)}} \tag{3}$$

where $P(t)$ - perimeter, $A(t)$ - area of region. To estimate of smoke areas perimeter on each frame we use the median filtration procedure after localization of potential smoke areas on them. Rate of change of turbulence in the image is calculated using (4).

$$\frac{d\Omega(t)}{dt} = \frac{2P'(t)A(t) - P(t)A'(t)}{4\sqrt{\pi \cdot A^3(t)}} \tag{4}$$

where P represents the perimeter/edges of the remaining moving gray features and A represents the area of the region. As the complexity of a shape increases (i.e., the perimeter increase with respect to the area) the value associated with $\frac{d\Omega(t)}{dt}$ increases.

As a last step we investigate the cumulative sum of each area for each frame and the obtained $\frac{d\Omega(t)}{dt}$ in this frame and create a set of value intervals, which are valid for smoke.

3 Experimental Research

Experimental study is to test the developed system on a series of 20 videos, half of which contain smoke (1-10 rows in the table 1), another half - not (11-20 rows in the table 1). Testing was performed using color features only, motion

features only and two types together. In all cases we, first, allocate a region of potential smoke using color or moving features and divide it to separate areas. Then we apply turbulence function and transparency analysis on them to smoke recognition. Testing was conducted at the same optimal settings of the system and for each video sequence is determined errors of the first (false reject rate - FRR) and second kind (false accept rate - FAR).

Table 1. Results of experiments with optimal settings

№	FileName	Color + turbulence		Moving + turbulence		All features	
		FRR	**FAR**	**FRR**	**FAR**	**FRR**	**FAR**
1	Clip_23_Smoke	0,13	0,17	0,32	0,12	0	0
2	ModifiedClip_31	0,12	0,08	0,2	0,07	0.3	0
3	sBehindtheFence	0,22	0,42	0,09	0,22	0,03	0,05
4	sBtFence2	0,1	0,33	0,09	0,17	0	0,02
5	sEmptyR1	0,22	0,5	0,17	0,42	0,01	0,05
6	sEmptyR2	0,47	0,42	0,43	0,39	0,22	0,27
7	ShorterIsyamNight_Short	0,29	0,55	0,25	0,35	0,15	0,02
8	Smoke_Manavgat_Raw	0,01	0,70	0,01	0,03	0,01	0
9	sParkingLot	0,26	0,55	0,21	0,27	0,11	0,02
10	sWasteBasket	0,07	0,28	0,05	0,07	0.01	0
11	Astak IP Cam	0	0,07	0	0,05	0	0
12	AXIS 212 PTZ	0	0,11	0	0,08	0	0
13	CCTV	0	0,17	0	0,13	0	0,02
14	Motion Test	0	0,22	0	0,17	0	0
15	my IPcamera 360	0	0,33	0	0,26	0	0
16	RVi-127	0	0,3	0	0,21	0	0
17	surveillance Camera	0	0,22	0	0,13	0	0
18	Ubiquiti airCam	0	0,18	0	0,14	0	0
19	VeSta VC-6320 IR	0	0,19	0	0,17	0	0
20	Vivotek IP8332	0	0,13	0	0,14	0	0

For video set from table 1 average values of FAR and FRR are 0.04 and 0.02 respectively. The maximum error value is 0.27. This is due to the rather complicated scenes and specificity of smoke in the "complex" video sequences. The especial interesting video is the video sequence *ShorterIsyamNight*, which in addition to the smoke contain the object - a man in a black and gray clothes, appered in a front of camera and walked down the hall, away from the viewpoint. The background color is mostly gray and adds additional complexity for recognition. As can be seen from the figure, a third-party object (in this case a man) is not recognized by the system as smoke. Smoke on the contrary, well-allocated by the system.

All video sequences are publicly available at the URL:
http://yadi.sk/d/q97BQ9v58WNRD.

4 Conclusion

In this paper we have shown an approach for early smoke detection. The main difference between smoke detection algorithms presented earlier is that in this work we analysis the speed of color changing of objects beyond the smoke area. This allows to detect smoke regions most robustly. Average false reject and false accept rate is about 5%. Disadvantage of this system - manual setting. The matter of future work is developing of autotune system.

References

1. Buchsbaum, G.: A spatial processor model for object color perception. J. Franklin Inst. 310(1), 1–26 (1980)
2. Catrakis, H.J., Dimotakis, P.E.: Shape Complexity in Turbulence. J. Phys. Rev. Lett. 80(5), 968–971 (1998)
3. Celik, T., Ozkaramanly, H., Demire, H.: Fire and Smoke Detection Without Sensors: Image Processing Approach. In: Proceedings of 15th European Signal Processing Conference EUSIPCO, pp. 1794–1798 (2007)
4. Chunyu, Y., Jun, F., Jinjun, W., Yongming, Z.: Video Fire Smoke Detection Using Motion and Color Features. J. Fire Technology 46(3), 651–663 (2010)
5. Fujiwara, N., Terada, K.: Extraction of a smoke region using fractal coding. In: IEEE International Symposium on Communications and Information Technology, ISCIT 2004, October 26-29, vol. 2, pp. 659–662 (2004)
6. Grech-Cini, H.J.: Smoke Detection. US Patent No. US6844818B2
7. Gudukbayb, U., Enis Cetin, A.: Computer vision based method for real-time fire and flame detection. Pattern Recognition Letters 27(1), 49–58 (2006)
8. Kopilovic, I., Vagvolgyi, B., Sziranyi, T.: Application of panoramic annular lens for motion analysis tasks: surveillance and smoke detection. In: Proceedings of 15th International Conference on Pattern Recognition, September 3-7, vol. 4, pp. 714–717 (2000)
9. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Contour Based Smoke Detection in Video Using Wavelets. In: Proceedings European Signal Processing Conference, EUSIPCO 2006 (2006)
10. Toreyin, B.U., Dedeoglu, Y., Cetin, A.E.: Wavelet based real-time smoke detection in video. In: Procedings of EUSIPCO 2005 (2005)
11. Vicente, J., Guillemant, P.: An image processing technique for automatically detecting forest fire. International Journal of Thermal Sciences 41(12), 1113–1120 (2002)

A New Nonlocal Maximum Likelihood Estimation Method for Denoising Magnetic Resonance Images

Jeny Rajan[1,2], Arnold J. den Dekker[3], Jaber Juntu[2], and Jan Sijbers[2]

[1] Department of Computer Science and Engineering, National Institute of Technology - Karnataka, Surathkal, India
[2] iMinds Vision Lab, Department of Physics, University of Antwerp, Belgium
[3] Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands

Abstract. Denoising of Magnetic Resonance images is important for proper visual analysis, accurate parameter estimation, and for further preprocessing of these images. Maximum Likelihood (ML) estimation methods were proved to be very effective in denoising Magnetic Resonance (MR) images. Among the ML based methods, the recently proposed Non Local Maximum Likelihood (NLML) approach gained much attention. In the NLML method, the samples for the ML estimation of the true underlying intensity are selected in a non local way based on the intensity similarity of the pixel neighborhoods. This similarity is generally measured using the Euclidean distance. A drawback of this approach is the usage of a fixed sample size for the ML estimation and, as a result, optimal results cannot be achieved because of over- or under-smoothing. In this work, we propose an NLML estimation method for denoising MR images in which the samples are selected in an adaptive way using the Kolmogorov-Smirnov (KS) similarity test. The method has been tested both on simulated and real data, showing its effectiveness.

Keywords: Image denoising, Kolmogorov-Smirnov test, MRI, Noise, Rice distribution.

1 Introduction

Denoising algorithms play an important role in the enhancement of MR images. Noise in MRI can be naturally minimized by averaging images after multiple acquisitions. This, however, may not be feasible in clinical and small animal MR imaging where there is an increasing need for speed [7]. Thus, post processing techniques to remove noise in the acquired data are important. Also, in time-sensitive acquisitions or studies with limited imaging time (diffusion MRI, functional MRI etc.), experiments cannot be repeated to do averaging. In the early days, many authors applied the conventional classical denoising techniques to denoise MRI [5]. These methods assumed the noise in the image to be Gaussian distributed. The major drawback of these methods is that the biasing effects

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 451–458, 2013.

of Rician noise, which characterizes magnitude MR images, were not taken into account. This bias increases with decreasing SNR. Later many methods were proposed to denoise MR images. Most of these methods exploited the second raw moment of the Rice distribution to reduce the bias in the denoised images[7,8]. That is, the image is denoised with the methods based on the Gaussian assumption and to reduce the bias, $2\sigma\_g^2$ is subtracted from the squared denoised image (where σ_g^2 is the variance of the noise in the complex image). However, in [14] it is shown that the sample size and SNR have a significant influence on the process of estimating the true underlying signal using this approach. ML methods were proved to be better than the aforementioned methods[9].

ML based denoising methods applied to magnitude MR images incorporate the Rice distribution to estimate the true underlying signal from a local neighborhood within which the signal is assumed to be constant. However, if this assumption is not met, such filtering will lead to blurred edges and loss of fine structures in the image. As a solution to the blurring issue of the local ML approach, the non local ML (NLML) estimation method [6] was proposed. In the NLML method, the samples for the ML estimation of the true underlying intensity are selected in a non local (NL) way based on the intensity similarity of the pixel neighborhoods. This similarity is generally measured using the Euclidean distance[6]. In that method, however, the number of NL pixels to be considered for ML estimation is fixed and is determined in a heuristic way. This fixed sample size can introduce under- or over-smoothing in the images. In this work, we propose a non local ML estimation method for denoising MR images in which the samples are selected in an adaptive way using the Kolmogorov-Smirnov similarity test.

This paper is organized as follows. Section 2 gives a short overview about the noise characteristics in MR images. In Section 3, the signal estimation using conventional NLML is discussed. Section 4 elaborates the proposed method. Section 5 presents the experimental results, comparative evaluation and discussion. Finally, conclusions and remarks are drawn in Section 6.

2 Noise Distribution in MR Images

The acquired complex valued MR signals in the k-space are characterized by a zero mean Gaussian probability density function (PDF). The k-space data are then (inverse) Fourier transformed to obtain the magnetization distribution. After the inverse Fourier transform, the noise distribution in the real and imaginary components will still be Gaussian distributed due to the linearity and the orthogonality of the Fourier transform. However, due to the subsequent transform to a magnitude image, the data will no longer be Gaussian but Rician distributed.

Let R and I represent the real and imaginary parts of the noisy complex MR data (corrupted by zero mean Gaussian, stationary noise with the standard deviation σ_g) with mean values μ_R and μ_I, respectively. Then the PDF of the reconstructed magnitude data M will be [12]

$$p_{\mathrm{M}}(M|A,\sigma_g) = \frac{M}{\sigma_g^2} e^{-\frac{M^2+A^2}{2\sigma_g^2}} I_0\left(\frac{AM}{\sigma_g^2}\right)\epsilon(M) \tag{1}$$

where $M = \sqrt{R^2+I^2}$, $A = \sqrt{\mu_R^2+\mu_I^2}$, $I_0(.)$ is the 0^{th} order modified Bessel function of the first kind and $\epsilon(.)$ is the Heaviside step function.

Proper estimation of the noise variance, σ_g^2, is important for effective denoising of MR images. Noise can be estimated from an MR image in a number of ways [15,16,1,11,10]. A survey of these methods is given in [2]. In this work, we used the object based methods proposed in [10] for the estimation of noise.

3 Signal Estimation Using NLML Method

Let $M_1, M_2, ..., M_n$ be n statistically independent observations within a region of constant signal intensity A. Then, the joint PDF of the observations is:

$$p(\{M_i\}|A,\sigma_g) = \prod_{i=1}^{n} \frac{M_i}{\sigma_g^2} e^{-\frac{M_i^2+A^2}{2\sigma_g^2}} I_0\left(\frac{AM_i}{\sigma_g^2}\right) \tag{2}$$

The ML estimator of A can now be computed by maximizing the likelihood function $L(A)$ or equivalently $\ln L(A)$, with respect to A [13]:

$$\ln L = \sum_{i=1}^{n} \ln\left(\frac{M_i}{\sigma_g^2}\right) - \sum_{i=1}^{n} \frac{M_i^2+A^2}{2\sigma_g^2} + \sum_{i=1}^{n} \ln I_0\left(\frac{AM_i}{\sigma_g^2}\right) \tag{3}$$

and

$$\widehat{A}_{\mathrm{ML}} = \arg\max_{A} (\ln L) \tag{4}$$

Nevertheless, to estimate the true underlying intensity $\widehat{A}_{ML}$ for each pixel i in the image using Eq. (4), samples $\{M_i\}$ with similar underlying intensity need to be selected. The straightforward approach to select samples $\{M_i\}$ is to select all pixels from a local neighborhood of i. However, it is clear that around edges and fine structures the assumption of uniform underlying intensity is violated, and, as a result, blurring will be introduced in the image. An alternate approach is to use non local (NL) pixels instead [6]. Conventionally, NL pixels are selected based on the intensity similarity of the pixel neighborhoods. If the neighborhoods of two pixels are similar, then their central pixels should have a similar meaning and thus similar gray values [19]. The similarity of the pixel neighborhoods can be computed by taking the intensity distance (Euclidian distance) between them [6]:

$$d_{i,j} = \|N_i - N_j\| \tag{5}$$

where $d_{i,j}$ is the intensity distance between the neighborhoods N_i and N_j of the pixels i and j. For each pixel i, the intensity distance d between i and all other non local pixels j as defined by Eq. (5), in the search window are measured. The first k pixels are then selected as $\{M_i\}$ after sorting the NL pixels in the

increasing order of the distance d for the maximum likelihood (ML) estimation. This method is more effective in terms of preserving edges and fine structures than the local ML estimation. However, one major concern with this approach is the proper selection of k. In the conventional NLML method[6], k is fixed and generally determined in a heuristic way. Low or high values of k can cause under or over smoothing. In the section below, we discuss an alternative approach to select the samples for ML estimation.

4 NLML Method Based on Kolmogorov-Smirnov (KS) Similarity Test

The research presented in this section is motivated by the study in [18]. In [18], it is observed that the PDF of an image S, resulting from the difference of two images R_1 and R_2 ($S = R1 - R2$) with the similar underlying intensity and corrupted with Rician noise, is symmetric and can be approximated with a Gaussian distribution of mean μ_s and standard deviation σ_s given by

$$\mu_s = 0, \quad \sigma_s = \sqrt{2}\sigma_r \tag{6}$$

where σ_r is the standard deviation of the Rice distribution. When the SNR increases, σ_s approaches $\sqrt{2}\sigma_g$ since σ_r approaches σ_g for high SNR.

In this work, we make use of the above knowledge, instead of using the Euclidian distance, to find the samples for estimating the true underlying intensity for each pixel in the image. For every pixel i, difference patches $D_{i,j}$ with other NL pixels j, based on the neighborhoods of i and j, are first created. If the neighborhoods N_i and N_j of the pixels i and j are similar, then the distribution of the difference patch $D_{i,j}$ can be approximated with a Gaussian distribution with the parameter values as given in Eq. (6).

To test the normality of the distribution of D, the KS test was employed. The KS test was used to compare the distribution of D to a standard normal distribution after standardizing D. The null hypothesis of the KS test is that the difference patches D are normally distributed, and this hypothesis is rejected if the p-value of the KS test statistic is below 0.05 (i.e., the result of the KS test is 1 if the test rejects the null hypothesis at the 5% significance level, 0 otherwise)[18]. So for each pixel i, the NL pixels with a p-value$>$0.05 are selected as the samples $\{M_i\}$ for the ML estimation.

The parameter, σ_r, can be computed directly from the neighborhood of the pixel of interest if the underlying intensity of the selected neighborhood region is constant. Also, σ_r will be always less than or equal to σ_g. The relationship between σ_r and σ_g is given as [10]:

$$\sigma_r = \frac{\sigma_g}{\sqrt{\varphi}} \tag{7}$$

where φ is a correction factor depending on the SNR and is in the range $[1; (2 - \pi/2)^{-1}]$. If the selected neighborhood contains pixels from more than one region,

then the estimation of σ_r will be difficult. However, if the estimated σ_r is greater than σ_g, we can assume that the neighborhood contains more than one region and in those cases, in Eq. (6), we replaced σ_r with σ_g. For the estimation of σ_g, we used the object based method proposed in [10].

5 Experiments and Results

To evaluate and compare the proposed NLML (NLML_{KS}) with the conventional NLML method, experiments were conducted on both synthetic and real MR images. For the experiments on the synthetic data, we used the standard MR image phantom of the brain obtained from the Brainweb database [4]. The phantom image was degraded with Rician noise for a wide range of noise levels and the denoising efficiency of both algorithms were evaluated based on the Peak Signal to Noise Ratio (PSNR), the mean Structural Similarity Index Matrix (SSIM)[17], Bhattacharrya coefficient (BC)[3] and the Mean Absolute Difference (MAD).

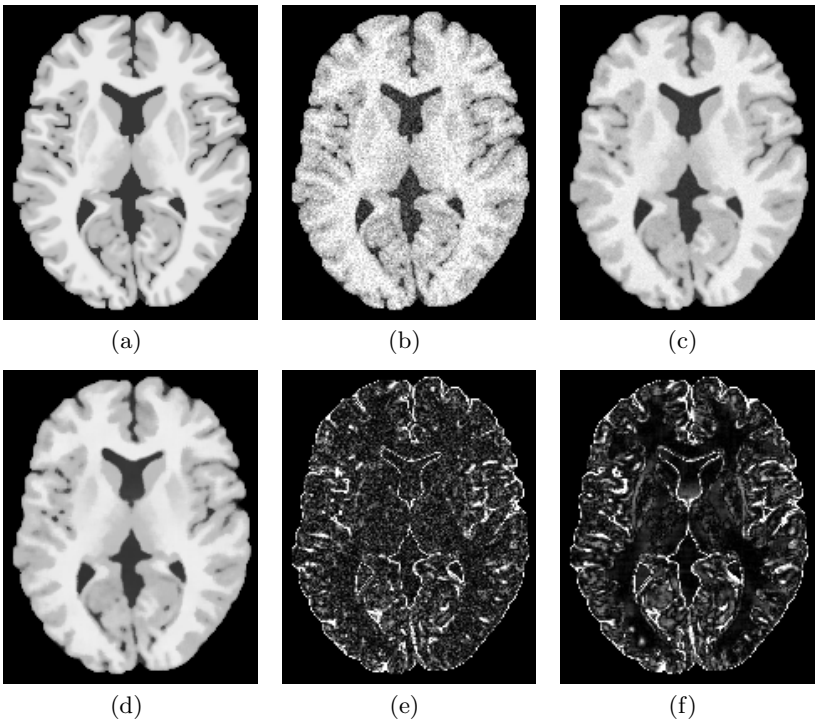


Fig. 1. Visual quality comparison: (a) Ground truth (b) Ground truth corrupted with Rician noise of $\sigma_g = 25$ (PSNR : 21.8, MSSIM : 0.5905) (c) denoised with conventional NLML method (PSNR : 28.4, MSSIM : 0.8626)(d) denoised with the proposed method (PSNR : 28.7, MSSIM : 0.9022)(e) NLML residuals (f) NLML_{KS} residuals. The residuals are scaled in the range 0-25.

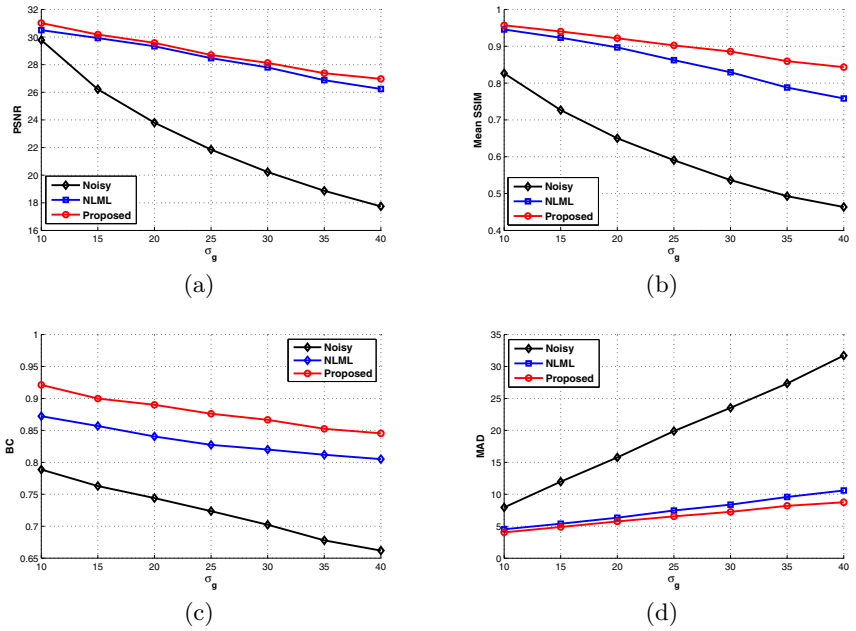
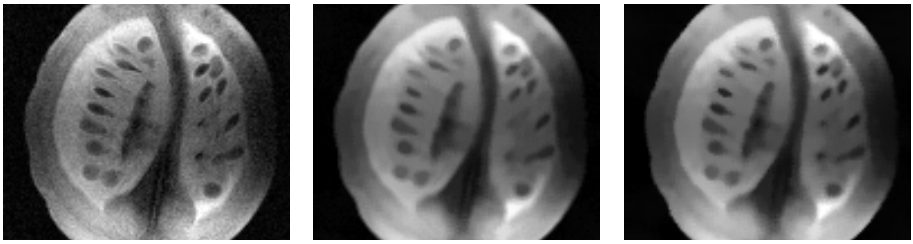


Fig. 2. Quantitative analysis of the proposed method with NLML based on (a) PSNR (b) SSIM (c) BC and (d) MAD for image corrupted with Rician noise of σ_g varying from 10 to 40

Fig. 1 shows the visual quality comparison of the image denoised with the NLML method and NLML$_{KS}$ method. This experiment was conducted on the brain image after corrupting the image by noise with $\sigma_g = 25$. It can be seen from the denoised images and from the corresponding residuals that the image denoised with the NLML$_{KS}$ method is closer to the ground truth than the image denoised with the conventional NLML. Fig. 2 shows the quantitative analysis of both methods in terms of PSNR, mean SSIM, BC and MAD. Both filters were executed with the following parameters: search window size $= 11 \times 11 \times 5$, neighborhood size $= 3 \times 3 \times 3$. For NLML, the sample size k was chosen as 25 (as recommended in [6]). It can be observed from the plots that the proposed NLML$_{KS}$ performs better than the conventional NLML based on the quality matrices used for the evaluation.

For the experiments on the real data, we acquired an MR image of a cherry tomato using an experimental MRI scanner (Bruker 7T scanner). The proposed NLML$_{KS}$ and the conventional NLML method were then applied on the acquired image. The results of this experiment is shown in Fig. 3. It can be seen from the images in Fig. 3 that the visual results are much better for the image denoised with the NLML$_{KS}$ in the sense that these images are less blurred than the images denoised with conventional NLML. This experiment on the real data set additionally indicates the effectiveness of the proposed method.



(a) (b) (c)

Fig. 3. Experiments on the MR image of cherry tomato. (a) Original image (b) denoised with NLML method (c) denoised with NLML_{KS}

6 Conclusion

We have proposed an NLML method based on the KS test to denoise the MR images. In the existing NLML method, the number of samples to be selected for ML estimation was based on heuristics. The proposed NLML method based on the KS test is statistically convincing and performs better (in terms of image quality) than the NLML method based on the Euclidian distance. Through the proposed KS test approach for the sample selection, the over and under smoothing caused by the conventional NLML can be reduced. Experiments have been carried out on simulated and real data sets. Quantitative analysis at various noise levels based on the similarity measures, PSNR, SSIM, BC and MAD shows that the proposed method is more effective than conventional NLML. Experiments were also performed on real MR images to prove the efficacy of the proposed method. However, it might be also possible to select the samples adaptively for the conventional NLML method based on a χ^2-test on the squared neighborhood differences. We are also working in that direction.

Acknowledgements. This work was financially supported by the Inter-University Attraction Poles Program of the Belgian Science Policy, and by the Fund for Scientific Research in Flanders.

References

1. Aja-Fernández, S., Alberola-López, C., Westin, C.: Noise and signal estimation in magnitude MRI and rician distributed images:a LMMSE approach. IEEE Trans. Imag. Proc. 17, 1383–1398 (2008)
2. Aja-Fernández, S., Tristán, A., Alberola-López, C.: Noise estimation in single and multiple coil magnetic resonance data based on statistical models. Magn. Reson. Imaging 27, 1397–1409 (2009)
3. Bhattacharyaa, A.: On the measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 35, 99–109 (1943)

4. Cocosco, C.A., Kollokian, V., Kwan, R.-S., Evans, A.C.: Brainweb: Online interface to a 3D MRI simulated brain database. NeuroImage 5(4), S425 (1997), http://www.bic.mni.mcgill.ca/brainweb/
5. Gerig, G., Kubler, O., Kikinis, R., Jolesz, F.A.: Nonlinear anisotropic filtering of MRI data. IEEE Trans. Med. Imag. 11(2), 221–232 (1992)
6. He, L., Greenshields, I.R.: A nonlocal maximum likelihood estimation method for rician noise reduction in MR images. IEEE Trans. Med. Imaging 28, 165–172 (2009)
7. Manjón, J.V., Carbonell-Caballero, J., Lull, J.J., García-Martí, G., Martí-Bonmatí, L., Robles, M.: MRI denoising using non local means. Medical Image Analysis 12, 514–523 (2008)
8. Manjón, J.V., Coupé, P., Martí-Bonmatíand, L., Collins, D.L., Robles, M.: Adaptive non local means denoising of MR images with spatially varying noise levels. J. Magn. Reson. Imaging 31, 192–203 (2010)
9. Rajan, J., Jeurissen, B., Verhoye, M., Van Audekerke, J., Sijbers, J.: Maximum likelihood estimation-based denoising of magnetic resonance images using restricted local neighborhoods. Physics in Medicine and Biology 56, 5221–5234 (2011)
10. Rajan, J., Poot, D., Juntu, J., Sijbers, J.: Noise measurement from magnitude MRI using local estimates of variance and skewness. Phys. Med. Biol. 55, N441–N449 (2010)
11. Rajan, J., Poot, D., Juntu, J., Sijbers, J.: Segmentation based noise variance estimation from background mri data. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010, Part I. LNCS, vol. 6111, pp. 62–70. Springer, Heidelberg (2010)
12. Rice, S.O.: Mathematical analysis of random noise. Bell. Syst. Tech. 23, 282–332 (1944)
13. Sijbers, J., den Dekker, A.J.: Maximum likelihood estimation of signal amplitude and noise variance from MR data. Magn. Reson. Med. 51(3), 586–594 (2004)
14. Sijbers, J., den Dekker, A.J., Scheunders, P., Van Dyck, D.: Maximum likelihood estimation of Rician distribution parameters. IEEE Trans. Med. Imag. 17(3), 357–361 (1998)
15. Sijbers, J., den Dekker, A.J., Verhoye, M., Van Audekerke, J., Van Dyck, D.: Estimation of noise from magnitude MR images. Magn. Reson. Imaging 16(1), 87–90 (1998)
16. Sijbers, J., Poot, D., den Dekker, A.J., Pintjens, W.: Automatic estimation of the noise variance from the histogram of a magnetic resonance image. Phys. Med. Biol. 52, 1335–1348 (2007)
17. Wang, Z., Bovik, A., Sheik, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. on Image Pocessing 13, 600–612 (2004)
18. Wink, A.M., Roerdink, B.T.M.: BOLD noise assumptions in fMRI. International Journal of Biomedical Imaging 2006, 1–11 (2006)
19. Zimmer, S., Didas, S., Weickert, J.: A rotationally invariant block matching strategy improving image denoising with non-local means. In: International Workshop on Local and Non-Local Approximation in Image Processing, Switzerland, pp. 135–142 (2008)

Object Shape Recognition from EEG Signals during Tactile and Visual Exploration

Anwesha Khasnobish[1], Amit Konar[2], D.N. Tibarewala[1], Saugat Bhattacharyya[2], and R. Janarthanan[3]

[1] School of Bioscience and Engineering, Jadavpur University, Kolkata
[2] Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata
[3] Department of Computer Science TJS Engineering College, Chennai, India
{anweshakhasno,biomed.ju,saugatbhattacharyya@gmail.com}, konaramit@yahoo.co.in,srmjana_73@yahoo.com

Abstract. Humans understand the world around us by visual and tactile exploration of the objects. The objective of this paper is to recognize the object-shapes from EEG signals while the subjects are exploring the same by visual and tactile means. The various object shapes are classified from electroencephalogram (EEG) signals that are stimulated by only tactile, only visual and by both means. EEG signals were acquired and analyzed from six electrodes, namely F3,F4,FC5,FC6,O1 and O2, where each pair of electrodes are located on frontal, somato-sensory and occipital region of the brain responsible for cognitive processing, tactile and visual perception. Mu-desynchronization in alpha and beta bands is used as the EEG modality for this purpose. Power spectral density (PSD) features are extracted and classified using support vector machine (SVM) classifiers in their corresponding object-shape classes. The results showed that object-shapes are best classified from EEG signals during only tactile exploration. The object shapes classified from EEG signals during only tactile exploration yielded highest mean classification accuracy of 88.34%. The average classification accuracy over all three object exploration modalities is 83.89%.

Keywords: Tactile perception, visual perception, object-shape recognition, electroencephalogram, power spectral density, support vector machine.

1 Introduction

Humans perceive their real world both by visual examination and manual exploration of the objects around them [1], [2], [3], [4]. This paper is targeted to recognize object shapes directly from EEG signals during visual and tactile exploration of the same objects. This study can find its applications in various areas of human computer interactions, tele-operation, and virtual reality. This study will also be helpful for blind (for whom visual examination is absent) as

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 459–464, 2013.

well as autistic patients (who are unable to express the information about the object perceived).

Electroencephalogram (EEG) provides an interesting means to understand brain functionality during visual and haptic percep-tion [5], [6], [7]. This paper considers EEG signal acquisition during only tactile exploration, only visual exploration and their joint consideration to recognize objects from EEG signals. The mu-desynchronization in alpha and beta band is out to be of im-portance of FFT analysis of the corresponding EEG signals during visual and haptic object exploration. EEG electrodes located particularly at frontal, somato-sensory, and occipital regions of the human brain following the traditional 10-20 EEG electrode placement system have been used for the present work. The electrode positions F3 and F4 refer to cognitive basis while FC5 and FC6 correspond to tactile and O1 and O2 correspond to visual basis of perception. Support vector machine (SVM) classifiers have been employed to recognize object-shapes from only tactile, only visual and the composite extracted power spectral density features. Experimental results indicate that only tactile perception provides the better basis of recognizing object shapes than only visual and composite modality.

The paper is divided in five sections. The selection of EEG modality and frequency bands and electrode localization is pre-sented in section 2. The undertaken methodology of feature extraction and classification of object shapes is explained in section 3. The section 4 depicts the experimental paradigm along with the results. Concluding remarks are included in section 5.

2 EEG Modalities for Visual/Tactile Perception

Performing FFT analysis on our data set, we chose the desynchronization of the mu rhythms for analysis of the tactile and visually stimulated event related potentials of EEG signals. It is observed that there occurs a decline in the EEG rhythms in mu-band on the onset of visual or tactile stimuli presented to the subjects. This is referred to as mu-ERD. Thus, we have considered these desynchronizations of mu signals as the EEG modality for analyzing tactile and visually stimulated events

In this work we are concerned with both visual and tactile perception for recognizing shapes. The visual perception originates from the occipital lobe; the tactile i.e. somatosensory perception is related to the somatosensory cortex. Along with these, cognitive processing and understanding is associated with frontal lobe of human brain. EEG signals from these areas are acquired from F3, F4, FC5, FC6, O1 and O2 electrodes which lie on the scalp over the frontal, somatosensory and occipital lobes respectively [8]. Since the dominant frequency bands for mu rhythm is mainly 8-13Hz and also 15-25Hz, we have thus considered EEG signals in alpha (8-13Hz) and central beta band (15-25Hz) for further analysis.

3 EEG Feature Extraction and Classification

For the EEG signals within frequency range 8-25Hz, power spectral density (PSD) (9) features are extract-ed for classification.The Welch approach was applied along with a Hamming window of length 64. The Welch method divides the times series data into overlapping segments, computing a modified periodogram of each segment and then the PSD estimates is averaged. Also the average power was obtained for each band.

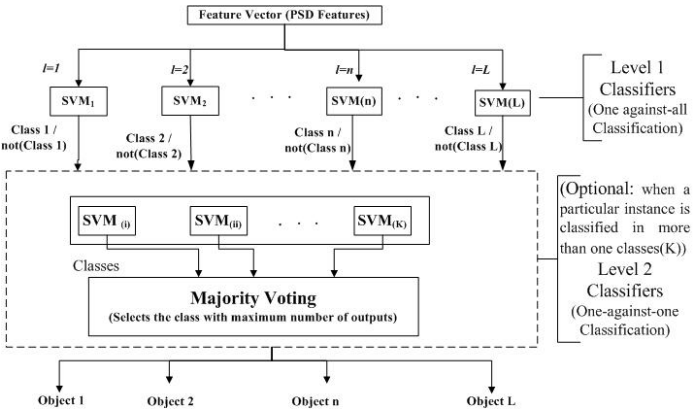


Fig. 1. One-against-all classification approach using linear SVM

The extracted PSD features from the EEG data were classified using support vector machine (SVM) [10], [11] into their respective object-shape classes.Using SVM we have classified the extracted PSD feature vectors in corresponding object classes (Fig.1). This is done for all the three experimental phases (i.e. only tactile exploration, only visual examination and by both means), which is explained in the section IV. Number of object classes considered in this study is l, (where $l = 1toL, L = 10$). For example, the instance of object 3(say) will be classified by SVM3 classifier, with output as Class 3 and the rest of the SVM classifiers will produce the output as not(class 3). The classification is performed in two stages. In the first level, the extracted PSD features are classified using L number SVM classifiers on one-against-all basis in their corresponding classes (Fig 2). Each SVM classifier, SVM_n, classified the feature vector in class n or not-n, where $n = 1toL$ i.e., according to the number of object classes. In ideal cases a particular instance will be classified in only one class (among all $l, l = 1toL$). If a particular instance after classification by Level *1* classifiers, is classified in more than one classes then the Level 2 classification is performed. For example, if a particular instant is classified in class 1, 3, 5 and 8. When next level of classification is performed to determine the actual class in which that particular instance belongs. In Level 2, $K = n_r^C$ number of SVM classifiers are employed, where n is the number of classes in which the particular instance is

classified (in the above example $n = 4$) and $r = 2$ since in this stage SVM will classify in one-against-one approach. The output clas-ses of these Level 2 SVM classifiers are fed in a majority voting system, which selects the class with maximum number of occurrences, which is the actual class in which that particular instance belongs. If there occurs no such class with more number of oc-currences than the other classes, then that particular instance is rejected. For this classifier system the cost (c) =1.00 and gamma=0.00 is set.

4 Experiments and Results

The experimental framework and the results obtained from them are illustrated in this section.

Framework. EEG signals are recorded from 25 subjects (17 male and 8 female) of age group 243 years. We considered 10 objects of different shapes as shown in Fig. 3. Our object set consisted of regular geometrical shapes such as cone, cube, sphere, hemisphere, cylinder, prism, and hexagonal base cylinder. Two irregular shaped objects i.e. a lock and a mouse is also included.EEG was recorded using a 14 channel Emotive EEG Research edition SDK with sampling rate of 128Hz. Out of the 14 channels we have considered F3, F4, FC5, FC6, O1 and O2 for our analysis purpose.

Fig. 2. Objects considered for experiments: Ob1. Cone, Ob2. Cube, Ob3. Cylinder, Ob4. Sphere, Ob5. Prism, Ob6. Hemisphere, Ob7. Pyramid (square base), Ob8. Hexagonal base cylinder, Ob9. Lock, and Ob19. Mouse.

The EEG data were recorded during three experimental phases consisting of only tactile exploration, followed by only visual examination and exploring the object by both visual and tactile means. Audio-Visual cues were presented to direct the subject to explore the object in the sequential manner for 10 seconds duration. A blank screen of 30seconds duration is followed by a fixation cross, directing the subject to focus on the screen. During only tactile exploration phase, the fixation cross is followed by an instruction to close the subjects eye, then the subject was given an object in his palm to dynamically explore by touching for 10seconds. A beep sound at the end of ten seconds prompts the

user to stop exploration and open his/her eyes. Then again a blank screen for 5 seconds is followed by the fixation cross and the sequence is repeated for exploring all the 10 objects. During visual examination, the fixation cross is followed by an image of the object on the screen for 10 seconds. During object exploration by both tactile and visually, the subject is prompted to dynamically explore the object by touching at the same time he/she can visually examine the same. During each phase, each object is examined 10 times for 10 seconds. Thus during each exploration phase (i.e. either visual, or tactile or both), for each subject data were recorded for 100 trials (10 objects examined 10 times) of 10 seconds each. The sampling rate was 128 Hz, thus we have 6400 data points for each object.

Performance Analysis. 10- fold cross validation is performed to divide the PSD features (see section 3) into training and testing instances and fed in the SVM classifiers.

Table 1. Classification Accuracies (in %)

Objects	(a) Only tactile	(b) Only visual	(c) Both tactile and visual
Ob1	77.8	77.8	88.9
Ob2	77.8	88.9	88.9
Ob3	88.9	77.8	77.8
Ob4	88.9	66.7	66.7
Ob5	100.0	87.5	75.0
Ob6	87.5	87.5	87.5
Ob7	100.0	75.0	87.5
Ob8	87.5	87.5	87.5
Ob9	100.0	75.0	87.5
Ob10	75.0	87.5	75.0
Mean	88.3	81.1	82.2

The results depicted in Table.1, shoe that the only tactile exploration yielded highest mean accuracy of 88.34%. The average classification accuracy over all three object exploration modalities is 83.89%.

5 Conclusion

In this paper we presented a novel approach to recognize object shapes from EEG signals during only visual, only tactile and both tactile-visual exploration. The extracted PSD features are classified using two-level SVM classifiers in corresponding object-shape classes for the three experimental phases.

The results reveal that only tactile exploration provides subjects with every detail of the object, thus object shape recognition rate is the highest. It is also evident from the results that any of the three object-exploration modalities is sufficient to classify objects shapes from EEG signals. The visual perception does not carry information of the object shape for the hidden surfaces and edges, which contributes in lowest recognition rate while examining the objects only visually. Though in general we experience object recognition seems to be faster by only visual examination, but that is achieved through years of learning process in our brain. In case on exploring the objects by both visual and tactile means number of information increases to a large amount due to combination of both visual and tactual information, which results in increased misclassifications. Further study in this field will be performed considering more number of subjects and objects with new tools and techniques.

Acknowledgments. This study has been supported by University Grants Commission, India, University with Potential for Excellence Program (UGC-UPE) (Phase II) in Cognitive Science, Jadavpur University and Council of Scientific and Industrial Research (CSIR), India.

References

1. Cristino, F., Colan, L.I., Leek, C.E.: The appearance of shape in visula perception: eye movement patterns during recognition and reaching. In: 3rd International Conference on Appearance, UK, pp. 125–127 (2012)
2. Hoosiang, H.: Differential surface models for tactile perception of shape and online tracking of features. IEEE Trans System, Man and Cybernetics 18(2), 312–316 (1988)
3. Szubielsca, M.: Prior visual experience, and perception and memory of shape in people with total blindness. British J. Visual Impairment 29, 60–81 (2011)
4. Santhian, K.: Visual cortical activity during during tactile perception in sighted and visually deprived. J. Developmental Psychobiology 46(3), 279–286 (2005)
5. Mustafa, M., Lindemann, L., Magnor, M.: EEG Analysis of Implicit Human Visual Perception. In: ACM Human Factors in Computing Systems, CHI (2012)
6. Christopher, P., Taylor, J., Thut, G.: Brain activity underlying visual perception and attention as inferred from TMS-EEG: A review. J. Brain Stimulation 5(2), 125–129 (2012)
7. Grunwald, M., et al.: Theta power in the EEG of humans during ongoing processing in a haptic object recognition task. J. Cog. Brain Research 11, 33–37 (2001)
8. Teplan, M.: Fundamentals of EEG Measurement. J. Measurement Sc. Review 2(2) (2002)
9. Herman, P., Prasad, G., McGinnity, T.M., Coyle, D.: Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Trans. Neural Sys. Rehab Eng. 16(4), 317–326 (2008)
10. Gunn Steve, R.: Support Vector Machines For Classification and Regression. Technical report, University of Southampton (1998)
11. Burges Christopher, J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Kluwer Academic Publishers, Boston

Bone Contour Tracing in Digital X-ray Images Based on Adaptive Thresholding

Oishila Bandyopadhyay[1], Arindam Biswas[2],
Bhabatosh Chanda[3], and Bhargab B. Bhattacharya[3]

[1] Department of CSE, Camellia Institute of Technology, Kolkata
[2] Department of IT, Bengal Engineering and Science University, Howrah
[3] Center for Soft Computing Research, Indian Statistical Institute, Kolkata

Abstract. An orthopedic X-ray captures bone images along with surrounding flesh and muscle components. Segmentation of the bone component with a sharp contour is a challenging task as the bone and flesh regions often have pixels with overlapping intensity range. In this paper, we propose a new technique of contour extraction by integrating an entropy-based segmentation approach with adaptive thresholding. The method eliminates the shortcomings of earlier derivative or deformable model based approaches, and can be fully automated. Experiments with several digital X-ray images reveal encouraging results especially for long-bone X-ray images.

Keywords: Medical Imaging, Bone X-ray, Entropy, Standard Deviation, Adaptive Thresholding, Tracing.

1 Introduction

Contour detection in X-ray or ultrasound bone images is an important task in emerging health-care automation systems [1][2][3]. Machine identification of fractures first requires extraction of the exact contour of the concerned bone structure. However, segmentation of the bone part and its contour extraction is a hard problem in digital X-ray image analysis as the pixel values of the bone shadow appear inseparably with those of the flesh and muscle components. The conventional image analysis methods based on derivative techniques or deformable models [2] either fail to detect the bone contour correctly or fail to admit full automation. The presence of noise and low contrast between bone and flesh regions makes the problem even worse. In this paper, we have proposed a new technique by integrating an entropy-based segmentation method [10] with adaptive thresholding-based contour tracing. The novelty of the technique lies in the fact that it rectifies the false discontinuities of bone contour while retaining the real (uneven) discontinuities caused by fractures. Experiments on several long-bone digital X-ray images demonstrate the suitability of the proposed method for contour generation and fracture related abnormality analysis.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 465–473, 2013.

2 Contour Detection

Common edge detectors like Prewitt, Sobel and Canny edge detector measure the gray level transition of locally connected pixels [4]. Goshtasby and Turner used Laplacian of Gaussian function [5] in their work on segmentation of cardiac MR image [6]. Canny edge detector uses two levels of thresholding. Scepanovic et al. used this function in their work on fast algorithm for probabilistic bone edge detection [7].

Deformable models such as active contour [8], active shape [9], and level set method [4] have also been used for medical image segmentation [4]. These methods are contour-based instead of region-based. An inherent weakness of these approaches is that they typically require good initialization of the model contour. If the model is poorly initialized, these approaches can be easily affected by noise and extraneous edges in the image, resulting in incorrect segmentation of the regions.

Atlas based or model based approaches overcome the weakness of deformable models by roughly applying atlas to the image before applying the deformable model. For these approaches, a training set or a reference map is required.

3 Proposed Algorithm

In an X-ray image, the bone parts appear along with the surrounding tissues or muscles (i.e., flesh). In many regions of an X-ray image, the intensity range of pixels belonging to the bone region and that of its surrounding flesh region may overlap, which makes the segmentation inaccurate. The proposed algorithm segments the bone region of an X-ray image from its surrounding flesh using an entropy-standard deviation based segmentation method and then applies an adaptive thresholding based technique to generate the bone contour. This modification overcomes the shortcomings of an earlier method [10].

3.1 Bone Region Segmentation from an *X*-ray Image

The major challenge in segmentation of bone region in any X-ray image lies in identification and extraction of flesh to bone transition region. Overlapping intensity range of flesh and bone region restricts the use of pixel-based thresholding or edge-based approaches as they often fail to produce accurate results. In the proposed algorithm, we generate local entropy image from the input X-ray image [10]. Local entropy image clearly identifies the flesh-bone transition points in a X-ray image with bright bone region and relatively darker flesh region. Few X-ray images appear with bright flesh region resulting in overlapping flesh and bone intensity range. In such cases, the entropy image often fails to identify the flesh to bone transition correctly. To overcome this problem, we compute local standard deviation for each pixel and multiplied it with local entropy to facilitate bone image segmentation.

Entropy Computation and Construction of Entropy Image. The entropy value represents the measure of randomness associated with pixel values

belonging to a region. Following Shannon's definition of entropy and Pun's representation of image entropy, for an image with G gray level and probability of k^{th} gray level P_k, the image entropy can be represented as

$$E = \sum_{k=0}^{G-1} P_k log_2 \left(\frac{1}{P_k}\right) = -\sum_{k=0}^{G-1} P_k log_2 \left(P_k\right) \tag{1}$$

where $P_k = \frac{n_k}{M \times N}$ and n_k is the number of pixels with gray level k and $M \times N$ is the size of the image [11]. Local entropy represents the discrepancy between two probability distributions on the same event space [12]. It is related to the variance of gray levels in the neighborhood of a pixel. Local entropy divides the image into separate regions and then analyzes each region as a separate information source. If a small neighborhood window Ω_k of size $M_k \times N_k$ is defined within the image, then entropy of Ω_k can be represented as:

$$E\left(\Omega_k\right) = \sum_{j=0}^{G-1} P_j log_2 \left(\frac{1}{P_j}\right) \tag{2}$$

where $P_j = \frac{n_j}{M_k \times N_k}$ denotes the probability of gray level j in the neighborhood Ω_k, n_j is the number of pixels with gray level j in Ω_k and $E\left(\Omega_k\right)$, the local entropy of Ω_k. Yan et al. had applied local entropy based thresholding for transition region extraction [12]. It was shown that the local entropy is larger for heterogeneous regions and smaller for homogeneous neighborhoods. So a transition region will have larger local entropy values than those in the non-transition regions of the image. Kang et al.[12] used local entropy values to identify transition regions on region extraction from coronary angiogram .

Statistical analysis of neighborhood pixel intensity distribution can be performed by calculating the local standard deviation of total number of pixel values in 9×9 neighborhood of each pixel. Standard deviation of a function is calculated from variance of the function. The variance of a data set represents the dispersion of data values about its mean. If a data set X is $[a_1 a_2 a_3 .. a_n]$ and its mean value is b, then variance

$$V(X) = \left[\left(a_1 - b\right)^2 + \left(a_2 - b\right)^2 + \left(a_3 - b\right)^2 + \cdots + \left(a_n - b\right)^2\right] / \left(n - 1\right) \tag{3}$$

Standard deviation of the data set is represented by square root of variance, i.e., $S = \sqrt{V(X)}$. A low standard deviation indicates that data points tend to be very close to mean value whereas a high standard deviation indicates that data points are spread out over a large range of value.

In the X-ray images where a bright flesh region appears with overlapping flesh-to-bone transition, the entropy image fails to identify the transition region distinctly (Fig. 1(b)). It has been empirically observed that this problem can be resolved by taking the product of the matrices of local entropy and local standard deviation (Fig. 1(c)). In the proposed algorithm, we calculate the local entropy values (LE) and local standard deviation values (LS) over a (9×9) neighborhood for each pixel of the X-ray image. Then we construct an intermediate gray image

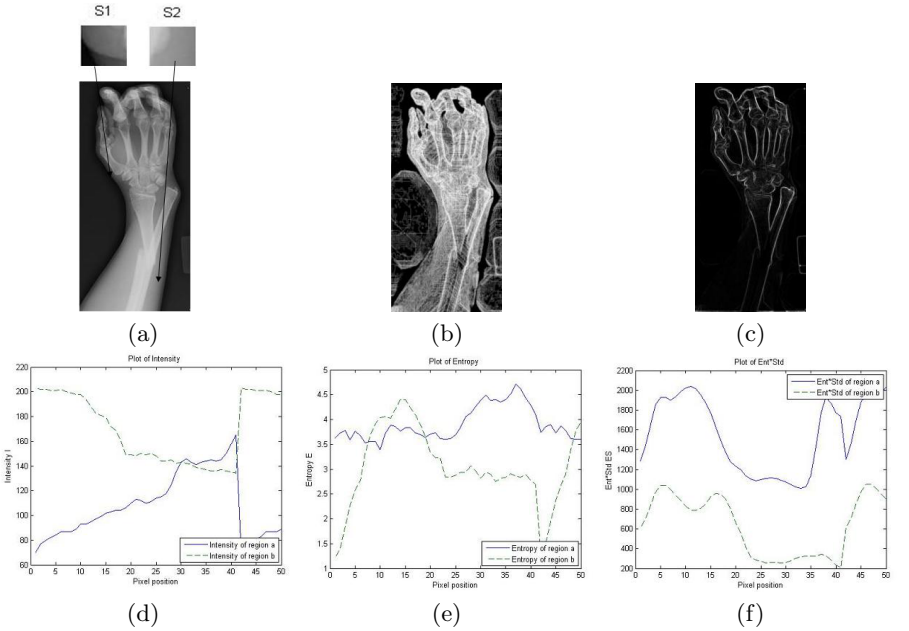


Fig. 1. (a) Wrist X-ray (b) Entropy Image (c) *Entropy* $\times$ *Standard deviation* image (d) Intensity Graph (s1 and s2) (e) Entropy graph (s1 and s2) (f)*Entropy* $\times$ *Standard deviation* graph (s1 and s2)

with each pixel value $I_i = (LE \times LS)$. In this process, the gray levels of the pixels at the transition region are enhanced compared to those of the neighbors.

In the X-ray image of wrist Fig. 1(a), it is found that the transition s1 of intensity from background to flesh varies between 40 to 160 and the intensity transition from bone to flesh s2 varies from 140 to 200. So there is a overlapping zone from 140 to 160 where the entropy change in two transition regions cannot reveal a distinct boundary for bone to flesh transition. In such a situation the product of local entropy and local standard deviation of input image generates distinct in-phase peaks for both s1 and s2 as shown in Fig. 1(d) and the resulting product image highlights the bone boundary distinctly [see Fig. 1(f)].

Contour Generation Using Adaptive Thresholding: The proposed algorithm takes the segmented bone image (entropy-standard-deviation image) as input to adaptive thresholding method [13] to generate the contour of bone part.

Threshold Computation: In the proposed algorithm, we traverse the segmented bone image J using a small window. For each pixel α of J, we construct the window consisting of its 8 neighboring pixels. The window is divided into four cells, top-left (C_1), top-right (C_2), bottom-left (C_3) and bottom-right (C_4). All these four cells incident on α. To determine whether a cell C_i has a portion of bone boundary in it, we use the adaptive thresholding approach [13]. In this approach, if max_i and min_i denote the respective maximum and minimum gray

level values in C_i, then the global maximum, at max_g, is defined as the maximum of all max_is for i = 1,2,3,4, corresponding to the four cells incident on α. Similarly, the global minimum min_g is defined as the minimum of all min_is for i = 1,2,3,4. Based on max_g and min_g, we define the local threshold value as $\tau_{loc} = \frac{1}{2}(max_g - min_g)$. Then, the adaptive value of the threshold τ_{adpt} is computed in accordance with the following equation

$$\tau_{adpt} = \gamma\tau_{loc} + (1-\gamma)(\tau_{loc} + \tau_{prev}) \tag{4}$$

where, $\gamma \epsilon [0,1]$ is the smoothing factor for the exponential moving average, and τ_{prev} is the adaptive threshold calculated for the previous point lying on the boundary of the cellular region. Realization of above equation is computationally simplified if we consider $\gamma = \frac{1}{2}$. For $\gamma = \frac{1}{2}$, equal weightage is given to the threshold of the previous inter-cell point and the local contrast, and hence the computation of τ_{adpt} can be effected by a right shift operation only. Once the adaptive threshold τ_{adpt} for a given common point is computed, then a cell C_i is said to contain a portion of bone boundary, provided $max_i - min_i \geq \tau_{adpt}$.
Initial Threshold Value: Initial value τ_0 is selected from the histogram of the intermediate entropy-standard-deviation image. As the bone boundary consists of a few pixels of intermediate image, we consider intensity value corresponding to 99.95 percentile in the intermediate image histogram as τ_0.
Contour Generation: The segmented bone image is traversed from five different directions like : (a) Top left to right bottom (Traverse-0) (Fig. 2(a)), (b) Top right to left bottom (Traverse-1) (Fig. 2(b)), (c) Top to bottom vertically (Traverse-2) (Fig. 2(c)), (d) Bottom right to top left (Traverse-3) (Fig. 2(d)), (e) Bottom left to top right (Traverse-4) (Fig. 2(e)). Based on the computed adaptive threshold value, the algorithm decides the direction of next movement and computes co-ordinates of the next pixel on bone boundary.

Computation of Direction: The algorithm decides on the selection of the next boundary pixel by checking the intensity values of its neighboring eight pixels. The direction of checking neighboring pixels may be clockwise (CW) or counterclockwise (CCW). This direction depends on the traversal mode like Traverse-0: CCW, Traverse-1: CW, Traverse-2: Top CW, Traverse-3: Bottom CW and Traverse-4: Bottom CCW. It depends on the position of bone region, i.e., the region of interest (ROI) in the respective contour. In the left (right) contour traversal, ROI should be on the right (left) side of the contour path. Our contour traversal algorithm checks the intensity values of neighboring pixels and selects the next pixel position whenever it encounters a pixel whose intensity value exceeds the adaptive threshold value of the present pixel. After selection of a new pixel position, the algorithm checks whether or not the pixel is already visited. If the pixel is found visited, then the algorithm starts searching from a new position; otherwise it adds the current pixel in the visited list and decides the direction for the next move. The use of adaptive thresholding rectifies most of the false discontinuities which appear in the outputs of other contour generation algorithms like Sobel's or Canny's method [see Fig. 3(k), Fig. 3(l)]. This happens because the adaptive thresholding algorithm works on a small pixel window and performs the thresh-

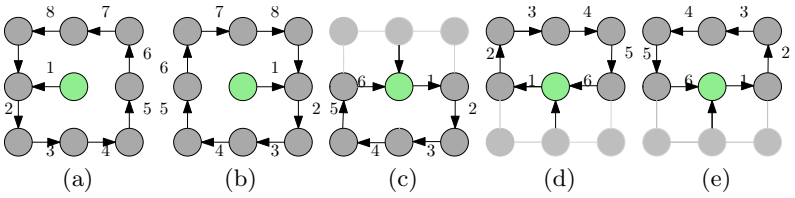


Fig. 2. Traversal Direction (a) Traverse-0, (b) Traverse-1, (c) Traverse-2, (d) Traverse-3 and (e) Traverse-4

olding with respect to the local maximum intensity. Hence, it produces the image contour correctly and also reduces the appearance of false edges.

Selection of New Start Point: In many cases, contour traversal ends with a pixel position which is already visited. In such situations, a new start point is selected based on the mode of traversal. For left-to-right (right-to-left) traversal, the search for a new start point is initiated from the left (right) end of next row. The process is repeated until no new start point is required. Finally, we connect the missing contour pixels in the boundary and generate the final boundary.

3.2 Experimental Results

We have used several X-ray images of long bone and generated the contour using the proposed algorithm. In first phase of the algorithm, we generate segmented bone image from input X-ray image (Fig. 3(b)). This segmented bone image ($I_i = LE \times LS$ image) becomes the input image for second phase of the algorithm. We traverse the segmented image and run the proposed adaptive thresholding based contour detection approach to generate the bone contour (Fig. 3(c)). The bone contour generated by this approach may have some short (false) discontinuities. The final pass of proposed algorithm connects the intermediate missing pixels of the bone contour (Fig. 3(d)).

We have run our algorithm on several input X-ray images and observed very good performance especially for long-bone X-ray images compared to Sobel and Canny edge detectors (see Fig. 3). In most of the images, the bone regions are clearly segmented and false discontinuities are removed. In some images, however, (e.g., the foot image of Fig. 3 (m)) in which the contrast is very poor at the transition region, complete reconstruction could not have been possible (Fig. 3 (p)). On the other hand, for a fractured bone image, it clearly shows the outline of the fracture in the contour as desired (Fig. 3 (s) - Fig. 3 (v)).

We quantify the performance of segmentation by using the SEQA measure [14]. The segmentation entropy $E(S)$ measures how much the image entropy has changed during segmentation. Better segmented regions are more homogeneous in nature and thus will exhibit lower region entropy. Table 1 shows that the $E(S)$ value of the bone contour generated by the proposed method is lower than that of the input X-ray image. Also, it produces improved results compared to Sobel's and Canny's not only in terms of $E(S)$ value, but also with respect to appearance of false discontinuities.

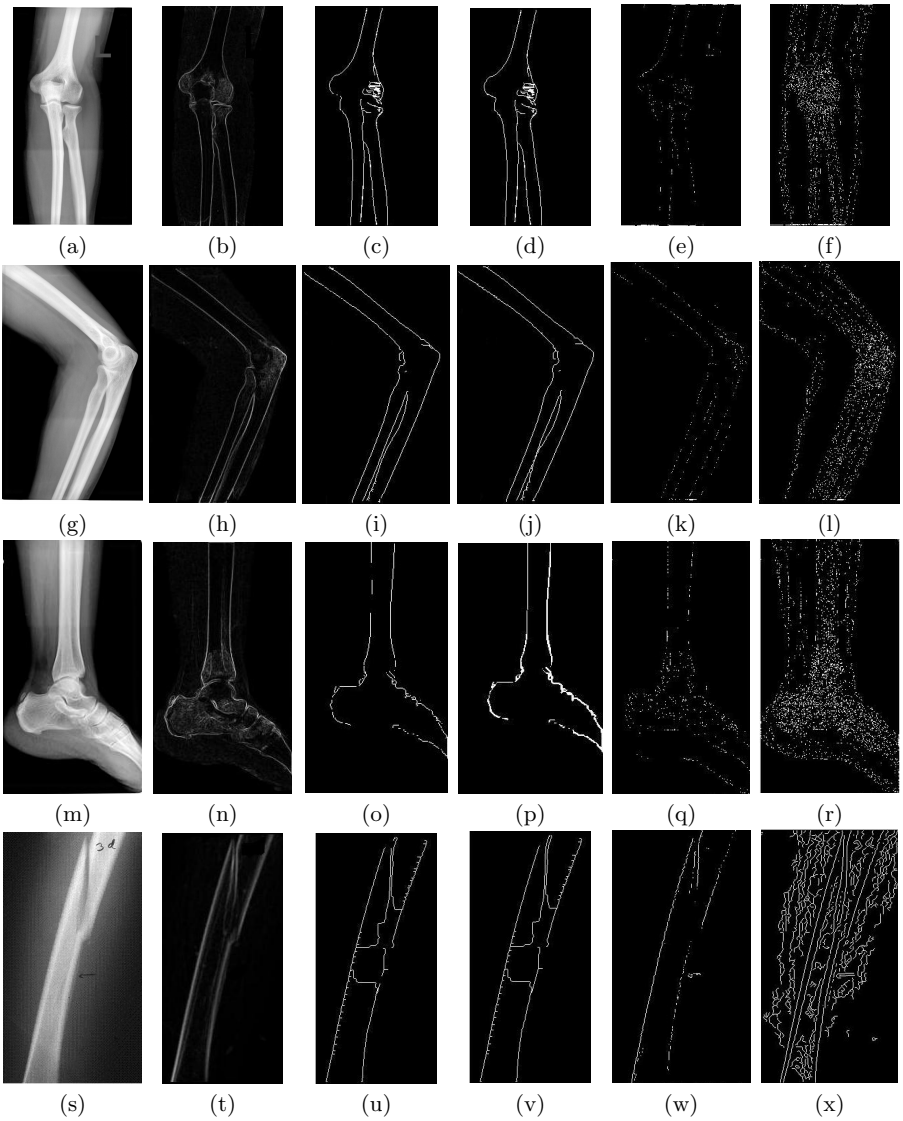


Fig. 3. Elbow X-ray (a) X-ray image (b) *Entropy* $\times$ *Standard deviation* image of (a) (c) Contour of (a) (d) Final Contour of (a) (e) Contour of (a) using Sobel (f) Contour of (a) using Canny (g) X-ray image (h) *Entropy* $\times$ *Standard deviation* image of (g) (i) Contour of (g) (j) Final Contour of (g) (k) Contour of (g) using Sobel (l) Contour of (g) using Canny (m) Leg X-ray image (n) *Entropy* $\times$ *Standard deviation* image of (m) (o) Contour of (m) (p) Final Contour of (m)(q) Contour of (m) using Sobel (r) Contour of (m) using Canny (s) Broken Tumor X-ray image (t) *Entropy* $\times$ *Standard deviation* image of (s) (u) Contour of (s) (v) Final Contour of (s) (w) Contour of (s) using Sobel (x) Contour of (s) using Canny

Table 1. PERFORMANCE MEASURE OF SEGMENTED X-RAY IMAGES

$E(S)$ of input image	$E(S)$(Sobel)	$E(S)$(Canny)	$E(S)$(proposed method)
5.09 (Fig. 3(a))	1.03	1.85	1.02
3.09 (Fig. 3(g))	1.05	1.65	1.04
4.51 (Fig. 3(m))	1.18	1.77	0.84
20.75 (Fig. 3(s))	0.66	1.63	0.63

4 Discussion and Conclusion

We have developed a new contour generation technique from gray-level bone X-ray images based on adaptive thresholding. This approach favorably overcomes the problem associated with the classical edge detectors like Canny's or Sobel's, particularly when the flesh boundaries appear as false bone edges. Contour generation approaches like deformable model based algorithms need good initialization of model contour. Our proposed method is fully automated and can generate the bone contour without any user intervention. Use of adaptive thresholding improves segmentation quality and expands the scope of its usefulness to more complex images. Bone contours thus generated from an X-ray image, can next be used for automatic fracture detection. Any non-uniformity present in the contour of a fractured bone image can also be identified by the algorithm while tracing the contour.

References

1. Eksi, Z., Dandil, E., Cakiroglu, M.: Computer-aided bone fracture detection. In: Proceedings of Signal Processing and Communications Applications, pp. 1–4 (2012)
2. Donnelley, M., Knowles, G.: Automated bone fracture detection. In: Proceedinds of SPIE, Medical Imaging: Image Processing, vol. 5747, p. 955 (2005)
3. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A.J., Rohling, R.N., Guy, P.: Automatic bone localization and fracture detection from volumetric ultrasound images using 3-d local phase features. Ultrasound Med. Biol. (1), 128–144 (2012)
4. Pham, D.L., Xu, C., Prince, J.L.: Current methods in medical image segmentation. Annual Review of Biomedical Engineering 2, 315–337 (2000)
5. Feng, D.: Segmentation of Bone Structure in X-ray Images. PhD thesis, School of Computing, National University of Singapure (2006)
6. Goshtasby, A., Turner, D.A.: Segmentation of cardiac cine mr images for extraction of right and left ventricular chambers. IEEE Transactions on Medical Imaging, 56–64 (1995)
7. Scepanovic, D., Kirshtein, J., Jain, A.K., Taylor, R.H.: Fast algorithm for probabilistic bone edge detection (FAPBED), vol. 5747 (2005)
8. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Computer Vision, 321–331 (1987)
9. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: The use of active shape models for locating structures in medical images. In: Barrett, H.H., Gmitro, A.F. (eds.) IPMI 1993. LNCS, vol. 687, pp. 33–47. Springer, Heidelberg (1993)

10. Bandyopadhyay, O., Chanda, B., Bhattacharya, B.B.: Entropy-based automatic segmentation of bones in digital x-ray images. In: Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K. (eds.) PReMI 2011. LNCS, vol. 6744, pp. 122–129. Springer, Heidelberg (2011)
11. Yang, J., Staib, L.H., Duncan, J.S.: Neighbourhood-constrained segmentation with level based 3-d deformable models. IEEE Transaction on Medical Imaging 23(8), 940–948 (2004)
12. Yan, C., Sang, N., Zhang, T.: Local entropy-based transition region extraction and thresholding. Pattern Recognition Letters, 2935–2941 (2003)
13. Biswas, A., Khara, S., Bhowmik, P., Bhattacharya, B.B.: Extraction of region of interest from face images using cellular analysis. Compute, 1–8 (2008)
14. Hao, J., Shen, Y., Xu, H., Zou, J.: A region entropy based objective evaluation method for image segmentation. In: IEEE International Conference on Instrumentation and Measurement Technology, pp. 373–377 (2009)

Content Based Image Retrieval of T2 Weighted Brain MR Images Similar to T1 Weighted Images

Abraham Varghese[1], Kannan Balakrishnan[2], Reji R. Varghese[3], and Joseph S. Paul[4]

[1] Adi Shankara Institute of Engineering and Technology, Kalady, Kerala
[2] Cochin University of Science and Technology, Cochin, Kerala
[3] Co-Operative Medical College, Cochin, Kerala
[4] Indian Institute of Information Technology and Management, TVM, Kerala

Abstract. Magnetic Resonance images play a crucial role in the diagnosis and management of the diseases of the brain. The MRI can acquire cross sectional images of our body, based on T1 and T2 relaxation of the tissues. As the information presented in these two images is often complimentary, both these images need to be compared for accurate clinical diagnosis. Hence automatic retrieval of similar slices of T1 weighted images from T2 weighted images or vice versa is of much value. In this paper T2-weighted (or T1 weighted) similar brain MR images within and across the subjects are retrieved using T1-weighted (or T2 weighted) as query images. The rotational and translational invariant form of Modified Local Binary Pattern is used to retrieve the similar anatomical structure MR images.

1 Introduction

The advantage of Magnetic Resonance Imaging (MRI) over other imaging modalities persuade radiologist to relies on MRI to study various brain related problems. Unlike Computer Tomography (CT), Single Photon Emission Computed Tomography (SPECT), and Positron Emission Tomography (PET), MRI operates at radio-frequency (RF) range; thus there is no ionizing radiation involved. Furthermore, MRI can generate excellent soft tissue contrast, and has the capability of producing images at any orientation. Moreover the information content present in MR images is extremely rich compared to other imaging modalities. After a magnetized spin system is perturbed from its equilibrium condition by an RF pulse, it will return to the equilibrium after a sufficient time once the perturbation is removed [1]. The recovery of longitudinal component is called T1 relaxation while the decay of transverse magnetization is called T2 relaxation. In the diagnosis process, radiologist depends on both T1 and T2 weighted images as the information presented in the two images is complementary. As the number of digital images produced is increasing in an alarming rate due to advancements in various imaging technologies, retrieving T2(or T1) images corresponds to T1(or T2) is challenging. A very popular approach of image retrieval

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 474–481, 2013.

is to first annotate the images by text and then use text-based database management systems (DBMS) to perform image retrieval [2,3,4,5,6]. But the difficulties such as subjectivity of human perception, amount of labour required etc faced by the manual annotation approach became more and more acute due to the emergence of large collection of images. To overcome these difficulties Content Based Image Retrieval (CBIR) was proposed. CBIR is any technology that helps to organize digital pictures by their visual content. The term CBIR appears in the literature in early 1990s in connection with retrieving images from a large database using color and shape features [7]. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features such as color, texture, shape etc which are extracted automatically from the images themselves [8,9,10,11]. Using the low level features, some of the CBIR systems such as QBIC[12], Photobook[13],Virage[14], VisualSEEK[15], Netra[16], SIMPLIcity[17] were developed.
In this paper similar anatomical T1 and T2 weighted images are retrieved using a texture descriptor Local Binary Pattern (LBP). The LBP is modified by incorporating local variance information into it and is termed as Modified Local binary Pattern(MOD-LBP).

2 Methodology

The overview of the image retrieval scheme is shown in Fig 1. The MOD-LBP image code is computed using a circular neighborhood with radius R and P neighborhood pixels. The MOD-LBP image is converted to polar form(r,θ) in order to compensate for rotation and translation in spatial description of the features by taking centroid as the origin of the image. The output image obtained is of size $\boldsymbol{N} \times \boldsymbol{N}$ with N points along the r-axis and N points along the θ-axis. The pixel value at the non-integer coordinate of the image is estimated using bilinear interpolation. The histogram of MOD-LBP is computed spatially, where the entries of each bin are indexed over annularly angularly partitioned regions. The pixel intensities are brought in the range [0, L], where L is a positive integer and normalized histogram of the image is taken as feature vectors for similarity computation. The images in the database are ranked based on the Bhattacharya distance between query and database images. The average rank and accuracy is calculated for a set of query images.

2.1 Pre processing

The image is converted to binary image by applying suitable thresholding. The components (objects) that have fewer than P(say) pixels are removed from the binary image, resulting only one connected component. The centroid of the binary image is computed using REGIONPROPS operation, which measures a set of properties for each connected component (object) in the binary image. Then resulting image is multiplied with original image for further feature extractions.

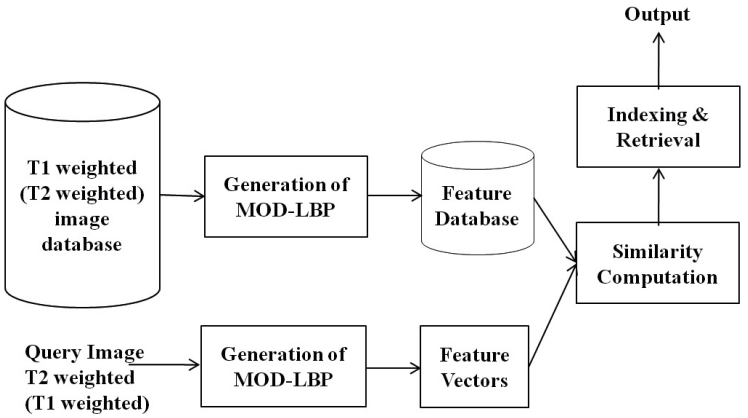


Fig. 1. Block diagram of Image retrieval scheme

2.2 LBP and MOD-LBP

The Local Binary Pattern (LBP) operator is a gray-scale invariant texture primitive, derived from a general definition of texture in a local neighborhood. Due to its discriminative power and computational simplicity, the LBP operator has become a highly popular approach in various computer vision applications, including visual inspection, image retrieval, remote sensing, biomedical image analysis, biometrics, motion analysis, environment modeling, and outdoor scene analysis, etc. LBP is formed by comparing gray value of center pixel (gc) with that of P neighborhood pixels in the local neighborhood [18,19].

$$LBP = \sum_{i=0}^{P-1} s(g_i - g_c)2^i \qquad (1)$$

The operator $LBP_{P,R}$, is derived based on a symmetric neighbor set of P members gi (i = 0, 1, P-1) within a circular radius of R. The signs of the differences Eq (1) in a neighborhood are interpreted as a P -bit binary number, resulting in 2^P distinct values for the LBP code. One way to eliminate the effect of rotation is to perform a bitwise shift operation on the binary pattern P-1 times and assign the LBP value that is the smallest, which is now referred to as $LBP^{ri}_{P,R}$.The above measure is invariant to monotonic gray level changes because just the sign of the differences only are considered instead of their exact values. In order to account for the changes in contrast and rotational invariance against grayscale shift with respect to the window taken, the definition of the LBP is modified as shown in Eq(2).

$$MOD - LBP(P,R) = \frac{1}{P}\sum_{i=0}^{P-1} s(g_i - g_c)(g_i - \mu)^2 \qquad (2)$$

where μ is the mean of the P circular neighborhood pixels and R is the radius . Bilinear interpolation is applied to the non-integer coordinate points on the

image in order to interpolate the pixels. The weights are assigned to the P-bit binary pattern based on the square of the gray scale difference between neighborhood pixel values and mean gray scale value. The resultant image is formed by assigning the decimal equivalent of the weighted binary number to the central pixel. As MOD-LBP is computed locally, it is invariant to monotonic gray level change, and it can even resist intra-image illumination variation as long as the absolute gray value differences are not much affected. Normalized histogram of MOD-LBP is taken as feature vector and Bhattacharya distance is used for similarity computation

3 Results

The slices (T1 and T2 weighted) used in this work, are acquired on a 1.5 Tesla, General Electric(GE) Signa HDxt MR Scanner from Pushpagiri Medical College Hospital, Tiruvalla, Kerala, India. Axial, 2D, 5mm thick slice images, with a slice gap of 1.5mm are acquired with the field-of-view (FOV) of range 220mm to 250mm. The T2 (TR/TE(eff.) of 3500-4500/ 85-105(eff.)ms) images are collected using Fast Spin Echo (FSE) sequences with a matrix size of 320 X 224 (Frequency X Phase) and a NEX (Averages) of 2. The T1 weighted images (TR/TE of 450-550/8-12ms) were acquired using standard spin echo (SE) sequence with a matrix size of 288 X 192 (Frequency X Phase) and a NEX(Averages) of 2.

We have categorized unregistered Brain MR images of different subjects into 4 levels, for independently evaluating the performance of the histogram based image retrieval method.
L1- the foramen magnum (The cerebellum with paranasal sinus is present)
L2- Above the fourth ventricle (Caudate nucleus, thalamus, basal ganglia are seen)
L3- mid ventricular section
L4- above the ventricle. In order to test the robustness of MOD-LBP with respect to intensity variations simulated bias fields from the BrainWeb MR simulator (20 and 40) are used [20]. These bias fields provide smooth variations of intensity across the image. As the MOD-LBP is calculated locally, intensity inhomogeniety has less sensitivity on it because the bias field in MRI is locally smooth. The degree of dissimilarity between MOD-LBP of original and the degraded images (100 images randomly chosen) are computed using the Bhattacharyya distance, $d = 1 - \sum_{i=0}^{L_1-1} \sqrt{p(i)q(i)}$ where p and q are normalized histograms with L1-bins. The mean dissimilarity scores fall in the range of [0 1], where 0 means that all original images and its degraded images are perfectly similar. Table 1 shows the effect of bias field in similarity computation between original image and degraded image using Bhattacharya distance. It is observed that the MOD-LBP computation helps to decrease dissimilarity between original and degraded image. Also as bias field increases the dissimilarity increases and MOD-LBP with window size 3 and 5 shows more or less same dissimilarity.Table 2 shows the time taken for MOD-LBP computation over different window size .

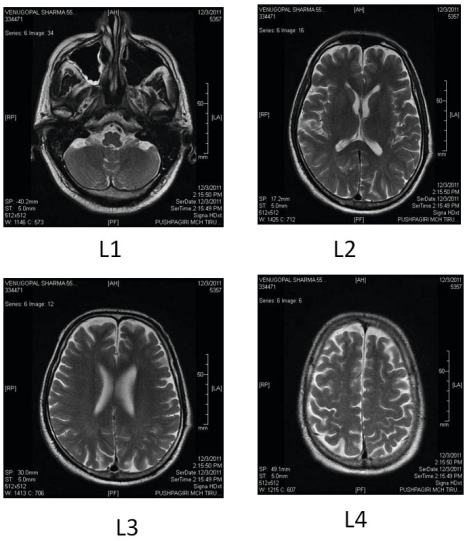


Fig. 2. Different classes of T2 weighted axial MR slices

3.1 Retrieval of T2 Weighted Images Similar to T1 Weighted Image across the Subjects

In order to test the performance of the retrieval system we had taken 40 T2 weighted images from L1, 32 images from L2, 32 images from L3 and 40 images from L4 as query image across the subjects and calculated average rank and accuracy of retrieving similar T1 weighted images based on the histogram of different LBP variants. An average rank, which shows the closeness of the system performance, is calculated using the formula,

$$Rank = \frac{1}{N_R}\left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R-1)}{2}\right) \quad (3)$$

N_R represents number of relevant images and Ri represents the rank at which the ith relevant image is retrieved [21]. The Accuracy of the retrieval system for a set of queries is also calculated using the formula,

$$Acc = \left(1 - \frac{No{:}of irrelevant images retrieved}{Total no{:}of images}\right) \times 100 \quad (4)$$

Table 3 illustrates the comparison of LBP and MOD-LBP in retrieving first 6 relevant images. It reveals that the MOD-LBP performs better in terms of average rank and accuracy for retrieving first 6 relevant images

3.2 Single Subject Performance

In order to study the retrieval performance evaluation in a single subject scenario, the most similar image corresponds to a given query from same subject data is retrieved. As our database consists of 20 slices of 20 subjects data, with four landmark slices per subject, there are 204 query images. For a query image i, the images in the database are ranked based on the distance value between

Table 1. The dissimilarity score between original and degraded image

Method	*WindowSize*	*BiasField*	
		20	40
Histogram		8.36×10^{-2}	12.26×10^{-2}
$MOD - LBP(8,1)$	3	4.93×10^{-2}	8.68×10^{-2}
$MOD - LBP(16,2)$	5	4.93×10^{-2}	8.83×10^{-2}

Table 2. Time taken for MOD-LBP computation over different window size

Operator	*Time*(S)
$MOD - LBP(8,1)$	$0.686s$
$MOD - LBP(16,2)$	$1.809s$

Table 3. The average rank and accuracy of retrieving first 6 T2 weighted images using T1 as query image

Class	*Method*	*Measure*	1	2	3	4	5	6
	LBP	*MeanRank*	2.07	3.10	3.98	4.93	5.87	6.93
*L*1		*Accuracy*	99.41	98.26	97.37	96.22	95.22	93.74
	MOD	*MeanRank*	2.93	3.43	4.02	4.55	4.99	5.47
	$-LBP(16,2)$	*Accuracy*	98.93	98.37	97.67	97.15	96.81	96.19
	LBP	*MeanRank*	3.80	5.43	8.02	10.85	13.71	16.46
*L*2		*Accuracy*	98.49	96.74	93.44	90.14	87.03	84.30
	MOD	*MeanRank*	2.93	3.97	4.71	5.43	6.40	7.26
	$-LBP(16,2)$	*Accuracy*	98.96	97.85	97.20	96.45	95.02	94.34
	LBP	*MeanRank*	3.67	5.10	5.98	7.08	8.16	9.30
*L*3		*Accuracy*	98.58	97.06	96.42	95.00	93.90	92.55
	MOD	*MeanRank*	1.80	2.33	2.84	3.38	3.92	4.42
	$-LBP(16,2)$	*Accuracy*	99.57	99.01	98.48	97.87	97.30	96.84
	LBP	*MeanRank*	1.13	1.27	1.33	1.37	1.43	1.51
*L*4		*Accuracy*	99.93	99.78	99.74	99.74	99.63	99.48
	MOD	*MeanRank*	1.0	1.07	1.11	1.15	1.23	1.29
	$-LBP(16,2)$	*Accuracy*	100	99.93	99.89	99.85	99.70	99.67

query and database image. The error in the retrieval of images correspond to ith query image is defined as $Err(i) = \sum_{j=1}^{K-1} |S_j - S_k|$ where Sj is the slice number at rank j, and k is the rank of the relevant slice. The retrieval error of the system is the mean of Err. Table 3 shows the mean retrieval error in retrieving relevant images. The results show that the retrieval error is minimum when the query and target images belong to the same class.

Table 4. Time taken for MOD-LBP computation over different window size .Mean retrieval error of retrieving relevant images a) Mean error of retrieving T2 weighted images using T2 weighted image as query b) Mean error of retrieving T1 weighted images using T1 weighted image as query c) Mean error of retrieving T1 weighted images using T2 weighted image as query.

	(a)	(b)	(c)
$LBP(8,1))$	11.72	8.25	22.68
$LBP(16,2)$	11.95	11.66	24.26
$MOD-LBP(8,1)$	5.66	3.13	16.9
$MOD-LBP(16,2)$	2.22	2.8	13.85

4 Conclusion

The paper illustrates a method to retrieve T2 weighted axial images from T1 weighted images and vice versa. The result reveals the robustness of local measure to intensity related problems. The method is applied to retrieve images from a single subject data as well as data across the subjects which would help the expert in the decision-diagnosis process. As the method is rotational and translational invariant, the costlier registration can be avoided and it is very easy to compute. The method can be extended to retrieve coronal and sagittal slices. An optimum performance can be achieved using proper relevance feed mechanism.

References

1. Einstein, A.: On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. Annalen der Physik 17, 549–560 (1905)
2. Bloch, F.: Nuclear induction. Physical Review 70, 460–474 (1946)
3. Chang, N.S., Fu, K.S.: A Relational Database System for Images. Technical Report TR-EE 79-28, Purdue University (May 1979)
4. Chang, N.S., Fu, K.S.: Query-by pictorial-example. IEEE Trans. on Software Engineering SE-6(6) (1980)
5. Chang, S.K.: Pictorial data-base systems. IEEE Computer (1981)
6. Chang, S.K., Hsu, A.: Image information systems: Where do we go from here? IEEE Trans. on Knowledge and Data Engineering 4(5) (1992)
7. Chang, S.K., Yan, C.W., Dimitroff, D.C., Arndt, T.: An intelligent image database system. IEEE Trans. Software Eng. 14(5) (1988)

8. Kato, T.: Database architecture for content-based image retrieval in Image Storage and Retrieval Systems. In: Jambardino, A.A., Niblack, W.R. (eds.) Proc. SPIE, vol. 1662, pp. 112–123 (1992)
9. Algrain, P., Zhang, H., Petkovic, D.: Content-based representation and retrieval of visual media, A review of the state-of-the-art. Multimed. Tools Appl. 3(3), 179–202 (1996)
10. Rui, Y., Huang, T., Chang, S.F.: Image retrieval: Current techniques, promising directions and open issues. J. Visual Commun. Image Represent. 10(1), 39–62 (1999)
11. Smeulders, A.W., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. 22(12), 1349–1380 (2000)
12. Snoek, C.G.M., Worring, M.: Multimodal video indexing: A review of the state-of-the-art. Multimed. Tools Appl. 25(1), 535 (2005)
13. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and effective querying by image content. J. Intell. Inf. Syst. 3(34), 231–262 (1994)
14. Pentland, A., Picard, R.W., Scaroff, S.: Photobook: Content-based manipulation for image databases. Int. J. Comput. Vision 18(3), 233–254 (1996)
15. Gupta, A., Jain, R.: Visual information retrieval. Commun. ACM 40(5), 70–79 (1997)
16. Smith, J.R., Chang, S.F.: VisualSeek: A fully automatic content based query system. In: Proceedings of the Fourth ACM International Conference on Multimedia, pp. 87–98 (1996)
17. Ma, W.Y., Manjunath, B.: Netra: A toolbox for navigating large image databases. In: Proceedings of the IEEE International Conference on Image Processing, pp. 568–571 (1997)
18. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: semantics-sensitive integrated matching for picture libraries. IEEE Trans. Pattern Anal. Mach. Intell. 23(9), 947–963 (2001)
19. Ojala, T., Pietikainen, M., Maeenpaa, T.: Multiresolution gray-scale androtation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)
20. Unay, D., Ekin, A., Jasinschi, R.: Medical image search and retrieval using local binary patterns and KLT feature points. In: Proc. Int. Conf. Image Process., pp. 997–1000 (2008)
21. `http://www.bic.mni.mcgill.ca/brainweb/`
22. Muller, H., Muller, W.: Automated Benchmarking in Content Based Image Retrieval, work supported by Swiss National Foundation for Scientific Research (grant no. 2000-052426.97)

Robust Classification of MR Brain Images Based on Multiscale Geometric Analysis

Sudeb Das and Malay Kumar Kundu

Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700108, India
to.sudeb@gmail.com, malay@isical.ac.in

Abstract. The widely used feature representation scheme for magnetic resonance (MR) image classification based on low-frequency subband (LFS) coefficients of wavelet transform (WT) is ineffective in presence of common MR imaging (MRI) artifacts (small rotation, low dynamic range etc.). The directional information present in the high-frequency subbands (HFSs) can be used to improve the performance. Moreover, little attention has been paid to the newly developed multiscale geometric analysis (MGA) tools (curvelet, contourlet, and ripplet etc.) in classifying brain MR images. In this paper, we compare various multiresolution analysis (MRA)/MGA transforms, such as traditional WT, curvelet, contourlet and ripplet, for brain MR image classification. Both the LFS and the high-frequency subbands (HFSs) are used to construct image representative feature vector invariant to common MRI artifacts. The investigations include the effect of different decomposition levels and filters on classification performance. By comparing results, we give the best candidate for classifying brain MR images in presence of common artifacts.

1 Introduction

Recently, magnetic resonance imaging (MRI) has emerged as one of the popular choice to study the human brain [1, 2]. However, because of the huge amount of imaging data, the existing manual methods of analysis and interpretation of brain images are tedious, time consuming, costly and subject to the experience of human observer. This necessitates the requirement of developing automated diagnosis tools to draw quicker and easier inferences from the MR images [3]. Various approaches of brain MR image classification are proposed by different researchers [3–8]. The general framework for classifying brain MR images mainly consists of three phases: (a) feature extraction, (b) feature reduction and (c) classification. The performances of the last two phases greatly depend on the effectiveness of the feature extraction phase. Most of the existing techniques use variants of wavelet transform (WT) for brain MR image's feature extraction. The coefficients of the low frequency subband (LFS) obtained after discrete WT (DWT) decomposition are used as the image representative feature vector. WT works effectively for brain MR images free from common MRI artifacts (small rotation, low dynamic range etc.), but fails in presence of these MRI artifacts.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 482–488, 2013.

WT and related classical MRA transforms suffer from several shortcomings: limited directionality, non-supportiveness to anisotropy etc. To solve these problems, some new multiscale geometric analysis (MGA) transforms such as curvelet (CVT), contourlet (CNT) and type-I ripplet (RT) are introduced [10–12]. These transforms provide superior results than traditional WT in many image processing applications. However, the effectiveness of these MGA transforms for brain MR image classification is not explored still. Our prime objective in this article is to construct an image representative feature vector invariant to these common MRI artifacts, and to compare the performance of different MRA/MGA tools using the same for brain MR image classification.

The rest of this paper is organized as follows. In Section 2, brief reviews of different MRA/MGA transforms are presented. In Section 3, we give the general classification framework for brain MR images and the novel feature extraction scheme. Section 4 presents details of numerical experiments and discussions on the results. Finally, conclusion is drawn in Section 5.

2 Multiresolution Image Decomposition

The MRA/MGA transforms investigated in this article are DWT, CVT, CNT and RT. In this section we briefly review these transforms.

2.1 Discrete Wavelet Transform

The advantage of wavelet is that it performs an MRA of a signal with localization in both time and frequency [9]. In addition to this, functions with discontinuities and functions with sharp spikes require fewer wavelet basis vectors in the wavelet domain than sine cosine basis vectors to achieve a comparable approximation. Discrete wavelet transform (DWT) can be implemented as a set of high-pass and low-pass filter banks. In standard dyadic wavelet decomposition, the output from the low-pass filter can subsequently be decomposed in the same way and the process continues to have finer resolution.

2.2 Curvelet Transform (CVT)

Traditional WT is unable to resolve 2-D singularities along arbitrarily shaped curves, and as a result cannot capture curves and edges of images effectively. To overcome this problem, Candes et al. proposed the CVT with the idea of representing a curve as a superposition of bases of various lengths and widths obeying the scaling law $width \approx length^2$ [10]. CVT uses a parabolic scaling law to achieve anisotropic directionality. A curvelet coefficient can be expressed as the inner product between an element $f \in L^2(R^2)$ and curvelet $\varphi_{j,l,k}$:

$$c(j,l,k) =< f, \varphi_{j,l,k} >= \int_{R^2} f(x)\overline{\varphi_{j,l,k}}dx = \frac{1}{2\pi^2}\int \widehat{f}(w)U_j(R_{\theta_l}w)e^{i<x_k^{(j,l)w}>}dw \quad (1)$$

where, $j = 0, 1, 2, ...$ is a scale parameter; $l = 0, 1, 2, 3, ...$ is an orientation parameter; and $k = (k_1, k_2), k_1, k_2 \, \epsilon \, \mathfrak{T}$ is a translation parameter. The $\varphi_j(x)$ is

defined by means of its Fourier transform $\widehat{\varphi_j}(\omega) = U_j(\omega)$, where U_j is frequency window defined in the polar coordinate system $U_j(r,\theta) = 2^{\frac{-3j}{4}} W(2^{-j}r)V(\frac{2\lfloor \frac{j}{2} \rfloor \theta}{2\pi})$ where, W and V are the radial and angular windows obeying the admissibility conditions, respectively. Curvelets at scale 2^{-j}, orientation θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 \cdot 2^j, k_2 \cdot 2^{\frac{-j}{2}})$ can be expressed as: $\varphi_{j,l,k}(x) = \varphi(R_{\theta_l}(x - x_k^{(j,l)}))$, where $\theta_l = 2\pi \cdot 2^{\lfloor \frac{j}{2} \rfloor} \cdot l$, with $l = 0, 1, ..., 0 \leq \theta_l < 2\pi$,$R_{\theta_l}$ is the rotation by θ_l radians. The actual implementation of the CVT is as follows: (i) apply the 2D FFT and obtain Fourier samples, (ii) form the product by multiplying the discrete localizing window $U_{j,l}$ for each scale j and angle l, (iii) wrap this product around the origin, (iv) apply the inverse 2D FFT to collect the discrete coefficients $c^D(j,l,k)$.

2.3 Contourlet Transform (CNT)

CNT gives a multiresolution, local and directional expansion of image using Pyramidal Directional Filter Bank (PDFB) [11]. The PDFB combines Laplacian pyramid (LP) which captures the point discontinuities, with a directional filter bank (DFB) which links these discontinuities into linear structures. LP is a multiscale decomposition of $L^2(R^2)$ into series of increasing resolution subspaces which are orthogonal complements of each other as follows: $L^2(R^2) = V_{j_0} \oplus (\bigoplus_{j=J_0}^{-\infty} W_j)$. An l-level DFB generates a local directional basis for $l^2(Z^2)$ that is composed of the impulse responses of the DFBs and their shifts. In CNT, the directional filter is applied to the detail subspace W_j. This results in a decomposition of W_j into 2^{l_j} subspaces at scale 2^j: $W_j = \bigoplus_{k=0}^{2^{l_j}-1} W_{j,k}^{l_j}$.

2.4 Ripplet Transform Type - I(RT)

To generalize the scaling law of the CVT and to find out which scaling law will be optimal for all types of boundaries Xu et al. proposed the RT. RT generalizes CVT by adding two parameters: support c and degree d [12]. CVT is a special case of RT with $c = 1$ and $d = 2$. In the frequency domain, the frequency response of ripplet function is in the form: $\widehat{\rho}_j(r,\omega) = \frac{1}{\sqrt{c}} a^{\frac{m+n}{2n}} W(2^{-j} \cdot r)V(\frac{1}{c} \cdot 2^{-\lfloor j\frac{m-n}{n} \rfloor} \cdot \omega - l)$ where, W and V are the 'radial' and 'angular' windows, respectively satisfying the admissibility conditions. a_j, $\overrightarrow{b}_k$ and θ_l represents the scale, position and rotation parameters, respectively and satisfy that $a_j = 2^{-j}$, $\overrightarrow{b}_k = [c \cdot 2^{-j} \cdot k_1, 2^{-j/d} \cdot k_2]^T$ and $\theta_l = \frac{2\Pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l$, where $\overrightarrow{k} = [k_1, k_2]^T$, and j, k_1, k_2, $l \in \mathbb{Z}$. $(\cdot)^T$ denotes the transpose of a vector. The 'wedge' corresponding to the ripplet function in the frequency domain is: $H_{j,l}(r,\theta) = \{2^j \leq |r| \leq 2^{2j}, |\theta - \frac{\pi}{c} \cdot 2^{-\lfloor j(1-1/d) \rfloor} \cdot l| \leq \frac{\pi}{2} 2^{-j}\}$. The DRT of an $M \times N$ image $f(n_1, n_2)$ is in the form of

$$R_{j,\overrightarrow{k},l} = \sum_{n_1=0}^{M-1} \sum_{n_2=0}^{N-1} f(n_1, n_2) \overline{\rho_{j,\overrightarrow{k},l}(n_1, n_2)} \tag{2}$$

where $R_{j,\overrightarrow{k},l}$ are the ripplet coefficients and $\overline{(.)}$ denotes the conjugate operator.

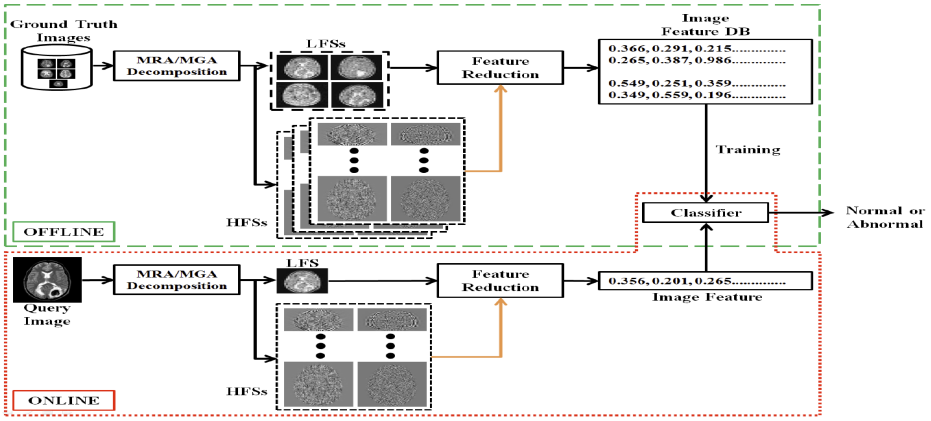


Fig. 1. Block diagram of the general framework for brain MRI classification

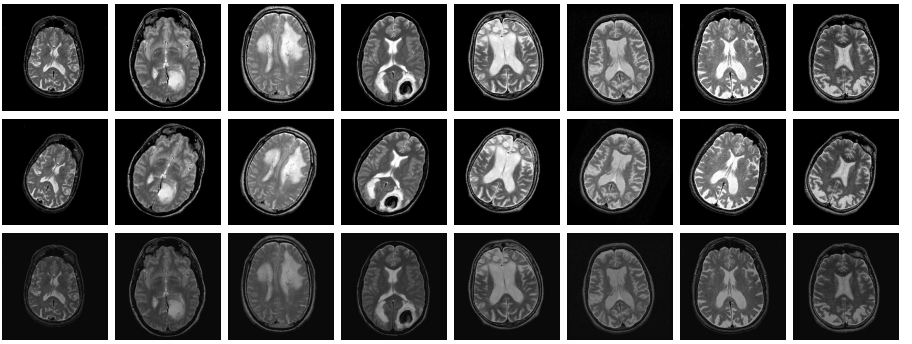
3 General Framework for Brain MRI Classification

The block diagram of the Fig. 1 shows the general framework for brain MRI classification. Different MRA/MGA transforms are used in the first phase for feature extraction. The high dimensionality of the feature vector is reduced by principal component analysis (PCA) during the second phase. In the third phase, various classification techniques are used to classify the brain MR image into normal and abnormal classes [3–8].

Small rotation, low dynamic range, noise etc. are the common artifacts present in today's MR images. The existing brain MR image classification schemes work efficiently for images free from common MRI artifacts, but their performance decrease in presence of these artifacts. The commonly used LFS for representing MR images is ineffective in capturing the subtle details and directional information present in the images. To tackle this problem, in this article we propose to use both the LFS and HFSs coefficients during feature extraction. After decomposing the images of the training set through a MRA/MGA transform, the subband's energy for each HFSs are calculated as follows:

$$ENG_s^d = \frac{1}{M \times L} \sum_{x=1}^{M} \sum_{y=1}^{L} |I_s^d(x, y)|^2 \tag{3}$$

where, I_s^d is the subband of size $M \times L$ at scale s and direction d, $s = 1, 2, ..., S$, $d = 1, 2, ..., D$. Then the coefficients of the LFSs and the HFSs energies' are passed to PCA for feature reduction, which results in the feature vector of dimension 10 (this value is chosen experimentally). The novel use of HFSs information for feature representation is indicated by "orange" colored lines in the block diagram of Fig. 1.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Sample brain MR images (row 1) with corresponding rotated (row 2) and dynamic range (row 3) modified versions: (a) normal, (b) glioma, (c) meningioma, (d) sarcoma, (e) Pick's disease, (f) Huntington's disease, (g) Alzheimer's disease, (h) Alzheimer's disease with visual agnosia

4 Experimental Results and Comparisons

The dataset used in the experiments consists of 66 T2-weighted MR brain images in axial plane, free from common MRI artifacts (downloaded from the website of Harvard Medical School URL:http://med.harvard.edu/AANLIB/). This dataset consists of 18 normal and 48 abnormal brain MR images. The abnormal brain MR images of the datasets includes images of the following diseases: glioma, meningioma, Alzheimer's disease, Alzheimer's disease plus visual agnosia, Pick's disease, sarcoma and Huntington's disease. The training set images are rotated through 5 different angles 5° to 25° in clockwise direction, and the dynamic ranges of the images are modified to 5 different levels (150,10) to (250,30), respectively. Therefore, the total number of images in the dataset used in the experiments is 726. To compare the performance of the DWT using different numbers of decomposition levels and wavelet bases, we consider 3 wavelet families: Daubechies (dbN, N = 1, 8), Coiflets (coifN, N = 1, 5), and Biorthogonal (biorN.N, N.N = 1.3, 6.8). For each wavelet basis, the number of decomposition levels from 1 to 4 is considered. For the other three transforms, 6 different pyramidal and directional filter combinations (from {'9/7', '5/3'} and {'9/7', '5/3', 'pkva'}) are considered with different number of decomposition levels. To classify the images k-nearest neighbor (k-NN) classifier is used with $k = 1$. To avoid overfitting problem, we integrate K-fold cross validation into our method ($K = 5$), which make the proposed scheme reliable and generalize to other independent datasets. Fig. 2, shows samples of the brain MR images used in the experiments.

We consider different number of principal components (PCs) to reduce the dimension of the feature vector. After extensive experiments we have found that 10 number of PCs (preserving $> 80\%$ of the total variance) is providing acceptable results. Therefore, in this article we compare the different MRA/MGA transforms considering only 10 PCs.

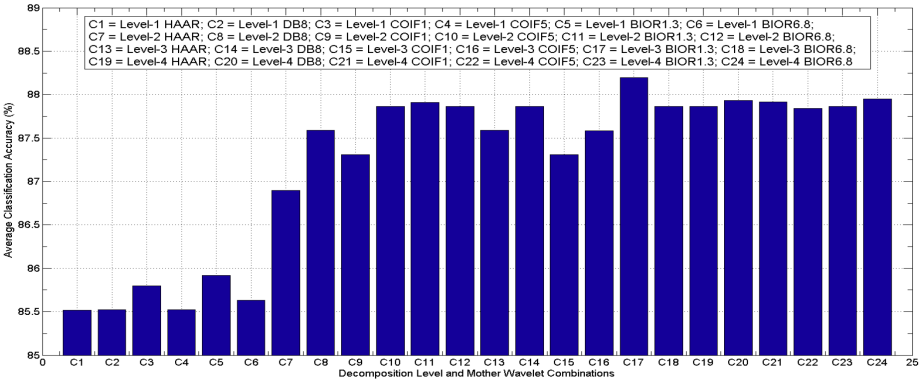


Fig. 3. Performance comparison of different wavelet combinations

Fig. 3 illustrates the performance comparison of WT with different decomposition levels and mother wavelets. For CNT, CVT and RT the best results are obtained for decomposition orientations {4, 8, 8}, {4, 4} and {2, 4}, respectively. The graph of Fig. 4 shows the results for CNT, CVT and RT with different filter combinations and the above mentioned best decomposition orientations. The

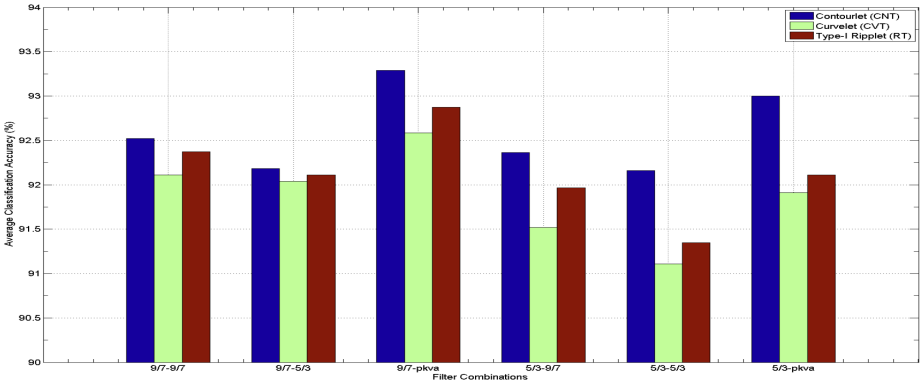


Fig. 4. Performance comparison of CNT, CVT and RT

graph of Fig. 3, shows that 'bior1.3' mother wavelet at decomposition level 3 provides the best performance ($\approx 88\%$) among different WT configurations. This is because of the high smoothness, small support, high vanishing moments, linear phase and near orthogonal properties of 'bior' wavelet compared to others. The performance decreases near about $10\% - 12\%$, when we only used LFS features for classification. From the graph of Fig. 4, it can be clearly seen that the filter combination '9/7' (pyramidal) and 'pkva' (directional) performs the best for CNT, CVT and RT. The '9/7' filters are linear phase and are close to being orthogonal and thus carries more subtle image information. Whereas, the ladder structure 'pkva' filters are more effective because of its superior edge direction localization property. Moreover, among these three transforms CNT performs

the best in terms of average classification accuracy of $\geq 93\%$. When we used only LFS features obtained from these three transforms for classification, the performance decreases near about $6\% - 7\%$.

5 Conclusions

Existing techniques for classifying brain MR images perform poorly in presence of common MRI artifacts. The directional information present in HFSs along with the LFS features prove effective in these scenarios. Moreover, recent advanced transforms perform much better than the conventional WT in classifying brain MR images. Comparative study of the performances of these transforms show that CNT performs superiorly considering both the classification accuracy and the dimension of the feature vector.

References

1. Westbrook, C.: Handbook of MRI Technique. John Wiley & Sons (2008)
2. Rombouts, S. A., Barkhof, F., Scheltens, P.: Clinical Applications of Functional Brain MRI. Oxford University Press (2007)
3. Zhang, Y., Wu, L.: An MR brain images classifier via principal component analysis and kernel support vector machine. Progress In Electromagnetics Research 130, 369–388 (2012)
4. Chaplot, S., Patnaik, L.M., Jagannathan, N.R.: Classification magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomedical Signal Processing and Control 1(1), 86–92 (2006)
5. Maitra, M., Chatterjee, A.: A Slantlet transform based intelligent system for magnetic resonance brain image classification. Biomedical Signal Processing and Control 1(4), 299–306 (2006)
6. Das, S., Chowdhury, M., Kundu, M.K.: Brain MR image classification using multiscale geometric analysis of ripplet. Progress In Electromagnetics Research 137, 1–17 (2013)
7. El-Dahshan, E.-S.A., Hosny, T., Salem, A.-B.M.: Hybrid intelligent techniques for MRI brain images classification. Digital Signal Processing 20(2), 433–441 (2010)
8. Zhang, Y., Dong, Z., Wu, L., Wang, S.: A hybrid method for MRI brain image classification. Expert Systems with Applications 38(8), 10049–10053 (2011)
9. Mallat, S.: A wavelet tour of signal processing: the sparse way, 3rd edn. Academic Press (2008)
10. Candès, E.J., Donoho, D.L.: Curvelets: a surprisingly effective nonadaptive representation for objects with edges. In: Cohen, A., Rabut, C., Schumaker, L.L. (eds.) Curve and Surface Fitting, pp. 105–120. Vanderbilt University Press (2000)
11. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Transaction Image Processing 14(12), 2091–2106 (2005)
12. Xu, J., Yang, L., Wu, D.: Ripplet: A new transform for image processing. Journal of Visual Communication and Image Representation 21(7), 627–639 (2010)

A Copula Based Statistical Model for Text Extraction from Scene Images

Ranjit Ghoshal[1], Anandarup Roy[2], and Swapan K. Parui[2]

[1] St. Thomas' College of Engg. and Technology
Kolkata- 700023, India
{ranjit.ghoshal.stcet,roy.anandarup}@gmail.com
[2] CVPR Unit
Indian Statistical Institute, India
swapan@isical.ac.in

Abstract. This article proposes a scheme for automatic text extraction from scene images. The work is composed of two steps. In the first step, we apply a model based color segmentation procedure in the LCH color space. This step produces certain homogenous connected components (CCs) from the image. In the next step, these CCs are examined in order to identify possible text components. A number of features that distinguish between text and non-text components, are defined. Further, during learning, these features are considered independently and approximated using parametric distribution families. Finally, the joint distribution of the features are constructed using a multivariate Gaussian copula. Consequently, we obtain two copula based class distributions for the two classes (text and non-text). Afterwards, during testing, a CC belongs to the class that produces the highest class distribution probability. Our experiments are on the database of ICDAR 2003 Robust Reading Competition. The experimental results are satisfactory.

1 Introduction

Automatic recognition of text portions in a scene image is useful to blind and foreigners with language barrier. Such a recognition methodology should also employ an extraction of text portions from the scene images. Segmentation of such text portions have a crucial impact on document processing, content based image retrieval, robotics etc. There have been several studies on text segmentation in the last few years. Wu et al. [1] use a local threshold method to segment texts from gray image blocks containing texts. images. Recently, Jung et al. [2] employed a multi-layer perceptron classifier to discriminate between text and non-text pixels. Recently Ghoshal et al. [3] proposed a scheme based on analysis of connected components (CCs) for recognition of Bangla text from scene images through perspective correction. Also a few criteria for robust filtering of text components have been studied.

We first employ a statistical mixture model based clustering algorithm on the input color image. With the assumption that text portions are homogeneous in

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 489–494, 2013.

color and lightness, different clusters may contain text portions as different CCs. We further study these CCs and define various features that are used to distinguish between text and non-text components. We consider the text identification as a two class problem. Each class (i.e., text and non-text) is approximated by a copula based distribution whose margins are the individual feature distributions. So, here the features are allowed to follow different parametric families. In the testing phase, we compare the probability of each CC against these two classes. The CC is classified in the class that has higher probability. Here, we use the public database of ICDAR 2003 Robust Reading Competition.

2 Color Image Segmentation

The primary task in this section is segmentation of the input color image. For this task we consider LCH color space. The hue component is angular, whereas chroma and the lightness components are linear. Hue can be represented by a random variable $\Theta \in [0, 2\pi)$. Thus, the LCH space has a circular linear characteristics. The random variable Θ, here, is assumed to follow a wrapped Gaussian distribution [4]. To accommodate the circular-linear characteristics, a suitable modeling should employ a distribution that is wrapped in the angular dimension and non-wrapped in the linear dimensions. The semi-wrapped Gaussian distribution [5] serves this purpose well enough. Roy et al. [6] discussed a statistical mixture model (SWGMM) with semi-wrapped Gaussian distributions. In this paper, we apply SWGMM in order to segmentation of an input image.

3 Connected Component Analysis

The image segmentation produces a number of CCs with one cluster producing one or more CCs. These CCs include the possible text portions. So, we analyze these CCs to identify text portions. We assume that a single text component is homogeneous in terms of color and lightness. This assumption ensures that a single text component is not broken after clustering. Now, after segmentation, the text parts may make one single cluster. However, more generally, one cluster contains non-text components along with some text components. In order to separate them we proceed as follows. We first remove sufficiently small and large components. Further, we extract the following CC based features to distinguish between text and non-text portions.

AR: The aspect ratio $AR = (height/width)$ of a non-text component is either very small or very large compared to text components.

OBR: The object to background pixels ratio (OBR) is computed by taking the bounding box. Due to the elongated nature of texts, only a few object pixels fall inside text bounding box. This ratio is usually larger for non-texts.

ER: The text like patterns are usually elongated. We use a measure of elongatedness (ER) of a component defined by Roy et al. [7].

TH: Thickness (TH) of a CC is computed from Ghoshal et al. [3].

H: Usually text like patterns contain less number of holes than that of non-text patterns. Using the Euler number we calculate the number of holes (H) inside a component.

Combining these features, we construct the 5-dimensional feature vector $\boldsymbol{\Upsilon} = (AR, OBR, ER, TH, H)$ for a CC.

4 Text Identification Method

We manually group the CCs into two classes, the text class (C_{text}) and non-text class ($C_{non-text}$). We now intend to obtain the feature distributions for these two classes separately. In this regard, the most popular model is the multivariate Gaussian distribution. However, the features under consideration here may not follow a Gaussian distribution. Hence, a non-Gaussian multivariate distribution may be a suitable choice to approximate the distribution of the present features. However, a drawback of using such a multivariate distribution is that all the margins usually follow distributions from the same family. For example, all the margins are Gaussian in case of a multivariate Gaussian distribution. In real situations, it is possible that the individual features may follow different families of distributions, independently. A suitable multivariate distribution should allow the marginal distributions to follow different families of distributions. The known families of multivariate distributions (for example, Gaussian or Dirichlet distributions) do not have this characteristic. However, there is an alternative construction of a multivariate distribution. In multivariate statistics, the copula approach is often taken to model the dependence between two or more random variables. Concerning the bivariate case, the copula approach to dependence modeling was first stated in a theorem due to Sklar [8]. Let H be a joint distribution function with marginal distributions F and G. Then there exists a copula C such that for all $x, y \in [-\infty, \infty]$,

$$H(x, y) = C(F(x), G(y)). \tag{1}$$

Accordingly, the probability density function (pdf) of $H(x, y)$ is defined as $h(x, y) = \frac{\delta^2 H(x,y)}{\delta x \delta y}$. Here, $C(u, v)$ is a mapping $[0, 1] \times [0, 1] \rightarrow [0, 1]$, termed as copula in the sense that it couples the random variables X and Y. The advantage of copula is that using the knowledge of the margins only, one can construct the joint distribution, accommodating however complex form of dependence structure, on the basis of different types of copulas. Sklar's theorem, we can construct the multivariate distribution incorporating different margin families. Let us start by specifying the marginal distributions. We consider each feature of a class, independently. Then we fit a series of parametric distributions on each feature. The best fitted distribution that maximizes the likelihood value, is selected for that feature. Here, we use an archive of distributions consisting of Gaussian, Gamma, Log-Normal and generalized extreme value (abbreviated by gextreme) distributions. Note that we obtain these distributions for each of the two classes

(namely, text and non-text) separately. To combine individual feature distributions, we use the multivariate Gaussian Copula [8]. Let $\boldsymbol{u} = \{u_1, \ldots, u_d\}$ be a d-dimensional (here $d = 5$) vector consisting of the distribution functions of all the margins. Here u_i denotes the distribution function of the i^{th} feature. Then $\boldsymbol{u} \in [0,1]^d$ according to the probability integral transform. The d-dimensional Gaussian copula is defined over $[0,1]^d$. It is constructed from a multivariate Gaussian distribution over $\Re^d$ by using the probability integral transform. Given the correlation matrix $\Sigma \in \Re^{d \times d}$, the density of a Gaussian copula can be written as

$$c_\Sigma(\boldsymbol{u}) = \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix}^T (\Sigma^{-1} - \boldsymbol{I}) \begin{pmatrix} \Phi^{-1}(u_1) \\ \vdots \\ \Phi^{-1}(u_d) \end{pmatrix} \right), \tag{2}$$

where Φ^{-1} is the inverse cumulative distribution function of a standard Gaussian distribution and Φ_Σ is the joint cumulative distribution function of a multivariate Gaussian distribution with mean vector zero and covariance matrix Σ. $\boldsymbol{I}$ is the identity matrix. We now combine all the five distributions for a class (say, for C_{text} class) and define the class distribution as:

$$f(C_{text}) = c_\Sigma(u_1, \ldots, u_5) \prod_{i=1}^{5} f_i. \tag{3}$$

Here, f_i is the density of the i^{th} feature. In a similar fashion, the probability distribution function $f(C_{non-text})$ for the non-text class can be computed. Finally, we classify a component to text class if $f(C_{text}) > f(C_{non-text})$.

5 Results and Discussion

Let us now present the segmentation and text extraction results. Here, we use the public database of ICDAR 2003 Robust Reading Competition. For the experiments, we select 200 images, randomly from this database. Further, we apply the SWGMM for segmenting the images into a set of homogenous CCs. A few sample images with segmentation results are presented in Table 3. We observe that our segmentation approach could preserve the text like components. For the purpose of training, we manually label the CCs of the training images to construct C_{text} and $C_{non-text}$. Our training set has 4000 samples of text components and 10000 samples of non-text components. We first perform a five-fold cross validation over the training data to assess the performance of our model. We get the cross validation accuracy 66.15% for C_{text} and 68.33% for $C_{non-text}$. Later, we obtain the distributions for the text and non-text classes. The individual features and their corresponding distributions are presented in Table 1. Note that most of the time we obtain the generalized extreme value family as a suitable distribution. In Table 1, we elaborate the parameter values of the distributions. In gextreme distribution, the first parameter (i.e., ξ) governs the tail behavior. The sub-family distributions corresponding to $\xi > 0$ and $\xi < 0$

are Fréchet and Weibull families respectively ($\xi = 0$ does not occur here). The other gextreme parameters μ and σ control the location and scale of the sub-family distribution. We also obtain Log-Normal distribution for the feature TH for the class C_{text}. The Log-Normal parameters are α and β that control the location and the shape of the distribution respectively. Then, we estimate the

Table 1. Parametric distribution correspond to individual features

Feature	C_{text}	$C_{non-text}$
AR	gextreme $(\xi = 0.1812, \sigma = 0.4743, \mu = 1.1071)$	gextreme $(\xi = 0.5747, \sigma = 0.5119, \mu = 0.5397)$
OBR	gextreme $(\xi = 0.3046, \sigma = 0.5304, \mu = 0.7791)$	gextreme $(\xi = 0.4825, \sigma = 0.8545, \mu = 0.9742)$
ER	gextreme $(\xi = -0.0781, \sigma = 2.6221, \mu = 9.9923)$	gextreme $(\xi = 0.2138, \sigma = 2.5187, \mu = 8.5799)$
TH	Log-Normal $(\alpha = 3.4281, \beta = 0.5264)$	gextreme $(\xi = 0.5102, \sigma = 7.5015, \mu = 10.9475)$
H	gextreme $(\xi = 4.2755, \sigma = 0.0242, \mu = 0.0057)$	gextreme $(\xi = 4.8188, \sigma = 0.1585, \mu = 0.0329)$

Gaussian copula parameter, i.e., the correlation matrix Σ. We plug in the previously estimated marginal distribution into the copula to obtain a maximum likelihood estimation of the correlation matrix. In Table 2, we present the correlation matrices obtained for the text and the non-text classes. Finally, after

Table 2. Correlation matrices of the Gaussian copulas for (a) C_{text} and (b) $C_{non-text}$

1	-0.19	-0.32	-0.37	-0.30
-0.19	1	-0.13	0.25	0.04
-0.32	-0.13	1	0.63	0.43
-0.37	0.25	0.63	1	0.32
-0.30	0.04	0.43	0.32	1

(a)

1	0.05	-0.09	0.18	0.01
0.05	1	0.20	0.41	0.12
-0.09	0.20	1	0.49	0.47
0.18	0.41	0.49	1	0.46
0.01	0.12	0.47	0.46	1

(b)

obtaining the distribution we now provide the test samples. We put each CC against the two classes and assign the component in class C for which $f(C)$ (Eq. 3) is larger. Some of the images that are the extracted text components (third row) are shown in Table 3.

We observe from Table 3 that often some non text components are included in the text class. On the other hand, some text components are still missing. However, during segmentation, some of the text components are not separated from the non-text portions. Such text portions are essentially not included in the extracted text components. We obtain satisfactory text extraction results.

Table 3. Some images (first row), the corresponding segmentation results (second row) and extracted text components (third row)

6 Summary and Future Scope

This article provides an automatic extraction of text entities from scene images. It is based on color image segmentation followed by extraction of several features that lead towards identification of text components. One primary objective is to obtain statistical models for feature distribution. We observe that different families of distributions can better approximate the distributions of individual features. However, it is quite possible that correlation is present among the features. So, we incorporate correlation among different features using multivariate Gaussian copula. However, the Gaussian copula omits the tail dependence and is thus less ambitious in modeling complex types of correlations. Our later work will address this issue and explore the use of other copula models.

References

1. Wu, V., Manmatha, R., Riseman, E.M.: Textfinder: An automatic system to detect and recognize text in images. IEEE Trans. Pattern Anal. Mach. Intell. 21(11), 1224–1229 (1999)
2. Jung, K., Kim, I.K., Kurata, T., Kourogi, M., Han, H.J.: Text scanner with text detection technology on image sequences. In: Proc. of Int. Conf. on Pattern Recognition, vol. 3, pp. 473–476 (2002)
3. Ghoshal, R., Roy, A., Parui, S.K.: Recognition of bangla text from scene images through perspective correction. In: Proc. ICIIP, pp. 385–390 (2011)
4. Mardia, K.V., Jupp, P.: Directional Statistics. John Wiley and Sons Ltd. (2000)
5. Bahlmann, C.: Directional features in online handwriting recognition. Pattern Recognition 39, 115–125 (2006)
6. Roy, A., Parui, S.K., Nandi, D., Roy, U.: Color image segmentation using a semi-wrapped gaussian mixture model. In: Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K. (eds.) PReMI 2011. LNCS, vol. 6744, pp. 148–153. Springer, Heidelberg (2011)
7. Roy, A., Parui, S.K., Paul, A., Roy, U.: A color based image segmentation and its application to text segmentation. In: Proc. of Ind. Conf. on Computer Vision, Graphics & Image Processing, pp. 313–319 (2008)
8. Nelsen, R.B.: An Introduction to Copulas. Springer (2006)

Query Expansion Using PRF-CBD Approach for Documents Retrieval

R. Rajendra Prasath and Sudeshna Sarkar

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur - 721 302, India
{rajendra,sudeshna}@cse.iitkgp.ernet.in

Abstract. Query Expansion has been widely used to improve the effectiveness of documents retrieval. In this work, we have attempted to identify additional terms for query expansion, from the initial set of documents retrieved for the original query, with the help of Clustering-by-Directions (CBD) algorithm proposed by Kaczmarek[2]. The CBD algorithm is based on a tag cloud of associated terms that are located in a radical arrangement and provides a clue to the direction of user intent in which search can be continued effectively. The output of the CBD approach gives rise to a set of terms in which we have considered top k terms for expanding the given query. The importance of these selected expansion terms is computed with respect to the number of terms in the radical of the selected directions. The experiments were conducted on FIRE 2012 adhoc data collection and we have performed monolingual documents retrieval in 3 major languages: Bengali, Hindi and English.

Keywords: Pseudo Relevance Feedback, Query Terms Expansion, Clustering-By-Directions, Query Terms Weighting.

1 Introduction

"Search for information" is one of the main activities for every human in his/her day to day life. It is difficult to imagine a world with the absence of search engines that strive to satisfy the information needs of different users on different aspects. When a user enters his information need in the form of a query composed of a set of keywords, a search engine is required to retrieve relevant information with respect to the actual intent of the query, not just based on weighted / non-weighted term matching. Thus, on many searches, the expectation of the people searching for specific information is not fulfilled merely by the subset of documents retrieved for the supplied set of keywords. In contrast, the actual user intent of the query has to be captured from the supplied keywords. So this tasks leads us to the problem of query expansion so as to have a deep understanding on the actual user intent.

In this work, we have considered the initial set of documents as Pseudo Relevance Feedback (PRF) and then applied the tag cloud based clustering by directions approach, for finding the terms for query expansion, that shows potential directions in which the search can be continued.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 495–500, 2013.

2 Query Expansion

There are many ways to refine ambiguous queries. Zhang *et* al. [6] used Open Directory Project to extract user interested topics and then used Wordnet to calculate semantic similarity of terms. In another work [1], different aspects of terms used in queries are analyzed to find meaningful groups of refinements. Inspired by the tag cloud based clustering-by-directions approach, we proposed documents retrieval system in which the PRF-CBD approach is employed for query terms expansion.

2.1 Pseudo Relevance Feedback

Documents retrieved by an IR system pertaining to a query may either be relevant or irrelevant. Relevance Feedback (RF) systems require users to mark the initial set of retrieved documents as relevant or irrelevant. IR system uses such manual feedback from users to further improve the performance of documents retrieval. To get improved retrieval performance without an extended interaction of users, Pseudo Relevance Feedback (PRF) [3] based methods have been proposed. The idea of PRF is to provide a mechanism for automatic local analysis that helps to find a better representative set of documents for the given query[3]. The PRF does documents retrieval to find an initial set of documents for a given query. The evidence captured from the content analysis of the initial set of retrieved documents is considered as PRF and used to identify a good set of representative terms for query expansion by means of the clustering-by-directions [2] algorithm.

2.2 Clustering-by-Directions Algorithm

The primary objective of the Clustering-By-Directions (CBD) algorithm[2] is neither to divide search results into clusters nor to show the related terms on the basis of some thesauri. It is applied to indicate the directions in which the search can be continued. To achieve this, the CBD algorithm first selects different directions, and afterward, it determines how the user can move in each direction. It is done regardless of whether there are subsets of web pages with the similar subject or not. The basic steps in the CBD algorithm are as follows:

1. Calculate vectors which represent documents and distances between these vectors;
2. Select different directions that meet the search needs in the space of information;
3. Assign documents to directions and select terms which represent directions;
4. Select top k terms as candidate terms for query expansion

Since documents are represented as vectors in the Vector Space Model(VSM), choosing the candidate direction is complicated due to a large number of dimensions in VSM. In order to choose representations serving as directions, the subset in which the sum of distances is the greatest is selected. For more details, readers may refer to [2].

3 Proposed IR System

The architecture of the proposed IR system with PRF-CBD approach is given in Figure. 1. The proposed IR system consists of two major components:

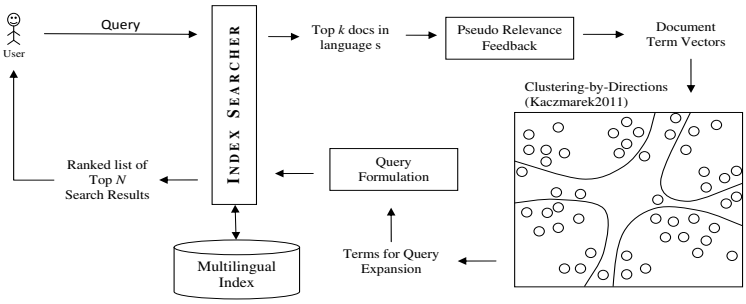


Fig. 1. The proposed PRF-CBD based IR system

(a) *Query Expansion:* We use the Clustering-by-Directions algorithm to choose the probable query terms and perform term weighting to formulate the expanded query
(b) *Ranking Documents:* We apply BM25 as the ranking function to compute the score of a document given the query.

3.1 Query Expansion

Initially we retrieve a set of representative documents pertaining to the given query. These initial set of documents are considered to be relevant. Then we apply CBD algorithm on the terms vectors of these documents and get a list of terms for each of the selected directions among d directions. Using these list of terms and their term statistics in top k documents, we formulate the weighted query, ensuring that at least one top ranking term from each direction exists in the reformulated query.

3.2 Ranking Documents

We have used Okapi BM25 [4,5] as our ranking function. BM25 retrieval function ranks a set of documents based on the query terms appearing in each document, regardless of the inter-relationship between the query terms within a document. Given a query Q, containing keywords $q_1, q_2, \cdots, q_n$, the BM25 score of a document D is computed as:

$$score(Q, D) = \sum_{i}^{n} idf(q_i) \cdot \frac{tf(q_i, D) \cdot (k_1 + 1)}{tf(q_i, D) + k_1 \cdot (1 - b + b \cdot \frac{|D|}{avgdoclength})}$$

where $tf(q_i, D)$ is the term frequency of q_i in the document D; $|D|$ is the length of the document D and *avgdoclength* is the average document length in the text

collection; k_1, $k_1 \in \{1.2, 2.0\}$ and b, $b = 0.75$ are parameters; and $idf(q_i)$ is the inverse document frequency of the query term q_i.

The inverse document frequency $idf(q_i)$ is computed as:

$$idf(q_i) = \log \frac{N - df(q_i) + 0.5}{df(q_i) + 0.5}$$

where N is the total number of documents and $df(q_i)$ is the number of documents containing the term q_i.

4 Experimental Results

4.1 Corpus and Topics

We have performed experiments with FIRE 2012 adhoc data collection on three languages: Bengali, Hindi and English. The coverage of documents in each collection is listed in Table. 1. We have used the query topics provided for each of these languages and also listed the details in Table. 1. We have submitted two runs (using title, desc fields) for each of the monolingual retrievals in Bengali, Hindi and English.

Table 1. FIRE 2012 adhoc dataset

Language	# documents	# terms	Topic IDs
Bengali	500,122	2,497,978	176-225
Hindi	331,599	1,164,526	176-225
English	392,577	1,427,986	176-225

4.2 Analysis

In this section, we illustrate our observations in detail. We report here only the runs of the monolingual documents retrieval in Bengali, Hindi and English. Since the relevance judgments of many Bengali documents are missing in evaluated pool, we are unable to get a clear picture from the partial results of Bengali monolingual runs. So we included only the partial results in the Precision - Recall plot. Figure. 2 shows the P-R curve of three monolingual retrievals: Bengali, Hindi and English, for the set of FIRE topics ranging from 176 to 225. In this experiment, we used title as the query and expanded it with PRF-CBD approach.

In the next experiment, we have used desc field as the query and expanded it using PRF-CBD approach. In Bengali Monolingual Retrieval, we observed that for the query id: 225 (equivalent query in English: "Satanic verses controversy"), PRF-CBD approach is able to present terms that capture the variety of news items on this topic. In Hindi monolingual retrieval with title field as the query, several queries, notably query ids: 179, 182, 184, 191, 199, 222, 223, 225 achieved a mean average precision greater than 0.7 and 60% of the query topics achieved map value of 0.5 or greater. Similar trends were observed when desc is used

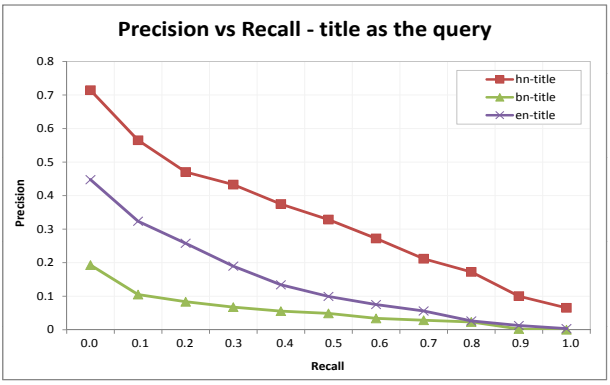


Fig. 2. P-R curve: title as the query in Bengali, Hindi and English monolingual retrieval

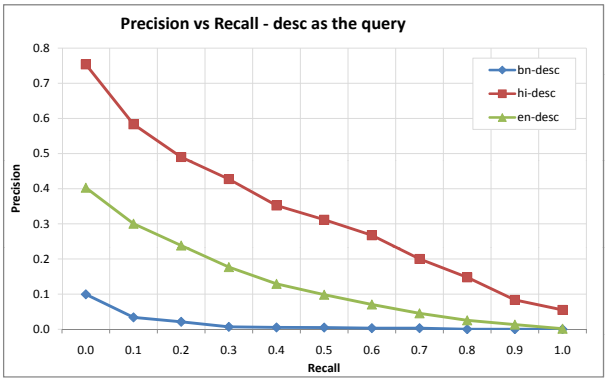


Fig. 3. P-R curve: desc as the query in Bengali, Hindi and English monolingual retrieval

as the query in Hindi Monolingual Retrieval. In English Monolingual Retrieval, almost 12 queries achieved map value of 0.25 and above with title as the query topic. The distribution of English monolingual retrieval with desc field resulted in a slightly degraded performance in terms of map score.

During the analysis of individual queries, we have observed query drifting issues with the follwing specific queries (with the quality of the retrieved results): *YSR Reddy death* - the retrieved results are mostly related to politics; *Chamunda Temple Stampede* - The retrieval system fetched news related to *dharshan* rather than stampede; and *Adarsh Housing Society Scam resignation* - The retrieved results are related to various scams and not related to Chief Minister Ashok Chavan's resignation. The system performed good in retrieving relevant documents for many queries like *countries adopting Euro*, *terrorists attack Indian parliament*, *Imran Khan Cancer Hospital* and so on. Figure. 4 shows the overall performance with precision at k documents ($p@k$) and Mean Averaged Precision (MAP) values of Bengali, Hindi and English monolingual retrieval on FIRE adhoc news corpus. In Hindi monolingual retrieval, when we used title as the query, several queries

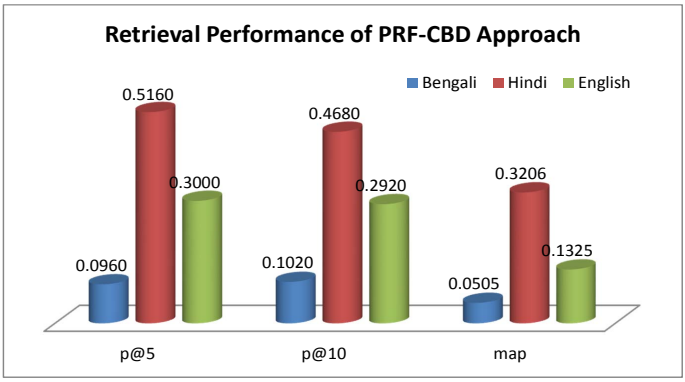


Fig. 4. Overall Performance of Bengali, Hindi and English monolingual retrieval

achieved the MAP value greater than 0.7 and in English monolingual retrieval, 12 queries achieved the map value of 0.25 and above with title used as the query. In the mean time, we noticed that no relevant document has been found at top 10 for 6 Hindi queries, 20 English queries and 18 Bengali queries.

5 Conclusion

In this work, we have proposed an approach to find a better representative terms for query expansion for effective document retrieval. We have tested the proposed monolingual documents retrieval system with PRF-CBD approach on FIRE 2012 adhoc data in three languages, viz., Bengali, Hindi and English. The proposed system performs better for many queries by finding a set of good representative terms. However, the derived results are sensitive to initial retrieved set of documents on which the CBD algorithm is applied.

References

1. Crabtree, D., Andreae, P., Gao, X.: Understanding query aspects with applications to interactive query expansion. In: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 691–695 (2007)
2. Kaczmarek, A.: Interactive query expansion with the use of clustering-by-directions algorithm. IEEE Trans. on Industrial Electronics 58(8), 3168–3173 (2011)
3. Manning, C.D., Raghavan, P., Schtze, H.: Introduction to Information Retrieval. Cambridge University Press, New York (2008)
4. Robertson, S., Zaragoza, H.: The probabilistic relevance framework: Bm25 and beyond. Found. Trends Inf. Retr. 3(4), 333–389 (2009)
5. Robertson, S.E., Walker, S.: Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In: Proc. of the 17th ACM SIGIR Conference on Research and Development in IR, SIGIR 1994, pp. 232–241. Springer-Verlag New York, Inc., New York (1994)
6. Zhang, B., Du, Y., Li, H., Wang, Y.: Query expansion based on topics. In: Fifth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 2, pp. 610–614 (2008)

Text Localization in Historical Document Images with Local Binary Patterns and Variance Models

Tapan Kumar Bhowmik[1] and Manika Kar[2]

[1] LITIS EA-4108, Université de Rouen, France
[2] Departamento de Engenharia Informática, Universidade do Porto, Portugal
tapan-kumar.bhowmik@univ-rouen.fr, manika.kar@fe.up.pt

Abstract. In this paper, we explore the utility of Local Binary Pattern (LBP) descriptors and variance measure towards the development of efficient techniques in order to segment a large collection of historical machine printed document pages. The result of segmentation will help us to organize the document pages in a structural format, which is useful in many applications like historical document access. In our experiments, three basic reference models namely background, text and image models are used to segment various non-text information together with the text. The method is tested on an archive of Portuguese historical documents and shows promising results.

1 Introduction

In recent years, the document analysis and recognition community has shown growing interests in processing of historical documents. These old documents often have historical and cultural significance and the aim is to scan them and create digital libraries [2], thereby offering continuous electronic access to this important part of the cultural heritage. To build such digital libraries, the first challenging task is to segment and classify the digitized documents into regions of text and image so that we can represent and store the scanned documents in a structural format. Representing documents in this way facilitate easy access to many non-textual information together with the text. During digitization process, in case of historical machine printed documents, OCR technology is used to convert text images into ASCII format for easy storage and retrieval. But, current OCR technology is largely restricted to find text printed against clean background, and cannot handle properly text printed against shaded or texture background, and/or embedded in images. However, more sophisticated text reading systems usually employ layout analysis schemes (page segmentation) for identifying text regions before applying OCR. But most of them [10, 8, 4, 5, 12] require fair background and the techniques are based on connected component analysis, which works on binarized version of input images. On the other hand, historical documents often suffer from several degradations and as a result it is hard to get clean binarized version of an input image. Therefore, we need a page segmentation scheme which works on natural background of the images without involving binarization.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 501–508, 2013.

In this paper, we describe such a text localization scheme based on multi-resolution analysis that we build as the first step of page layout system for historical documents. We assume that a document page has different texture regions due to different backgrounds, texts and images. So, we consider the page segmentation problem as a texture segmentation problem where segmentation process will partition the image into regions based on their texture [6, 3, 11, 7]. We explore the utility of various Local Binary Patterns (LBP), one of the most popular local image descriptor for texture analysis and use in our text localization scheme. In addition, variance measure is also used to localize the text. The proposed scheme is evaluated on an archive of Portuguese historical documents [1] and gives promising results.

2 Local Binary Pattern

Local Binary Pattern (LBP) was introduced by Ojala et. al. [9] for describing the local texture of gray scale images. The LBP operator describes each pixel (c) in an image with a certain binary pattern by calculating the difference of the gray values from a central pixel (g_c) around its circular neighborhood (with radius $R > 0$). If the difference of gray values between a neighboring pixel and the central pixel is greater than or equal to zero, the value is set to one, otherwise is set to zero (see in Figure 1(a)). Formally, the LBP operator is defined as follows:

$$LBP_{P,R}(c) = \{s(g_p - g_c)\}_{p=0}^{P-1} \simeq \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \text{ where}$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and $g_p(p = 0, 1, \ldots, P-1)$ correspond to the gray values of P equally spaced pixels on a circle of radius R, which forms a circularly symmetric neighbor set. If the coordinates of g_c are assumed to be $(0, 0)$, then the coordinates of g_p are calculated by $(Rsin(\frac{2\pi p}{P}), Rcos(\frac{2\pi p}{P}))$. Figure 1(b) illustrates circularly symmetric neighbor sets for various (P, R). The gray values of neighbors which do not fall exactly in the center of pixels are estimated by interpolation. Since, the difference of intensity is considered only while forming the binary pattern, so by definition the description is invariant against any monotonic transformation of the gray scale.

2.1 Rotation Invariance LBP

The basic $LBP_{P,R}$ operator produces 2^P different binary patterns that can be formed by the P pixels in the neighbor set. For example, $LBP_{8,1}$ operator can produce 256 unique binary patterns, which are not rotationally invariant. These are made rotation invariant by selecting the smallest value of $P-1$ bitwise shift operations on the binary pattern. Formally, the operator is defined as: $LBP^{ri}_{P,R}(c) = min\{ROR(LBP_{P,R}(c), i)|i = 0, 1, \ldots, P-1\}$, where $ROR(., i)$

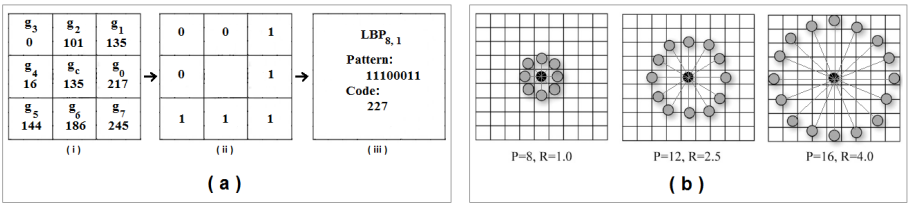


Fig. 1. (a) Example for calculating the LBP code. (b) Three circularly symmetric neighborhood sets (gray) around a central pixel (black) for different (P, R). If a point does not fall exactly on a pixel grid, the value is interpolated.

performs a circular bit-wise right shift on the P-bit binary pattern (say, $\{s(g_p - g_c)\}_{p=0}^{P-1}$) i times. For example, rotation invariant of $LBP_{8,1}(c) = 227$ in Figure 1(a) is $LBP_{8,1}^{ri}(c) = min\{227, 241, 248, 124, 62, 31, 143, 199\} = 31$. It is also be noted that there are 36 unique rotation invariant local binary patterns that can occur in the case of $LBP_{8,R}^{ri}(c)$ [9].

2.2 Uniform and Non-uniform LBP

From the practical experience, it was observed by Ojala et. al. [9] that certain binary patterns are fundamental properties of texture, providing the vast majority of patterns, sometimes over 90%, among all 3×3 patterns present in the observed textures. From this intuition, uniform and non-uniform patterns are defined. A pattern is considered as a uniform, symbolically called $LBP_{P,R}^{riu2}(c)$, if the number of transactions in the sequence between 0 and 1 are less than or equal to two. For example, 00000000_2 and 00000010_2 are the uniform patterns, as they have exactly zero and two 0/1 transitions respectively. Similarly, 10000010_2 is the non-uniform pattern because it has three 0/1 transitions. Non uniform patterns are considered the patterns that contain the main part of the noise of the images. Notice that a region with no transition is either a background or a flat region of the image. It is also be noted that, when $P = 8$, there exists only 9 uniform patterns out of 36 unique rotation invariant patterns [9].

2.3 Representing Image with LBP

For describing an image with LBP, each and every pixel is examined and found the distribution of the binary patterns for the whole image. It is fact that the different texture images have different distributions of binary patterns. This is done by calculating the histogram of binary patterns. The number of bins in the histogram depends on the choice of the operator. For example, the number of bins for $LBP_{8,R}(c)$ and $LBP_{8,R}^{ri}(c)$ are 256 and 36 respectively. It is observed that uniform patterns are present in distributions a lot more frequent than non-uniform patterns. Giving more weight to uniform patterns provide better discrimination. This is done by grouping all non-uniform patterns into single bin

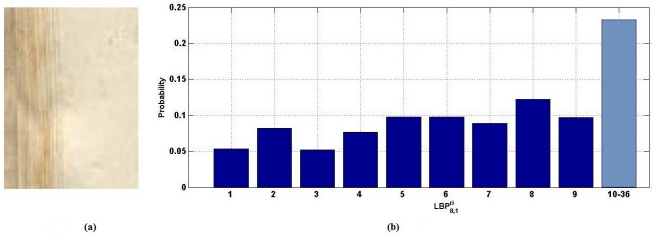


Fig. 2. Representation of an image with LBP. (a) image and (b) its corresponding $LBP_{8,1}^{riu2}$ histogram.

of the distribution. For example, in case of $LBP_{P,R}^{riu2}(c)$ operator the number of uniform pattern is 9 and non-uniform is $36 - 9 = 28$. But the histogram with number of bins 10, each uniform pattern is assigned to separate single bin and all the non-uniform patterns are assigned to another single bin (see in Figure 2).

3 Variance Measure

The variance is estimated in measure of standard deviation in local region of an image, namely window/block. Same LBP operators (described in previous section) are used to get a center pixel and its neighbor pixels (see in Figure 1(a)(i)). For a pixel (c) in a block image and a given operator $LBP_{P,R}(c)$, the standard deviation is measured by $\sigma(c) = [\frac{1}{P}\{(g_c - \mu(c))^2 + \sum_{p=0}^{P-1}(g_p - \mu(c))^2\}]^{\frac{1}{2}}$, where mean $\mu(c) = \frac{1}{P}[g_c + \sum_{p=0}^{P-1} g_p]$. If there is N number of center pixels (c) in a block image (w), the average variance measure is calculated as: $\sigma_{avg}(w) = \frac{1}{N}\sum_{c=0}^{N}\sigma(c)$.

4 Proposed Text Localization Approach

The input image is first divided into logical grids with fixed size disjoint window. For each window, LBP histogram feature is extracted. While building histogram feature, three LBP-operators $LBP_{P,R}^{ri2}(c), R = 1, 2, 3$ are used simultaneously due to preserving the multi-resolution property as well. For R equal to 1, 2 and 3, the ideal value of P is 8, 16 and 24 respectively. During the binary pattern calculation, only $P = 8$ equidistant points are considered. The remaining points (in case of $P = 16, 24$) help to find the gray values of 8 points. For example, when $P = 16$, the gray value (g_{p_i}) of a point p_i is considered the median gray value of three (2R - 1) consecutive points say, $p_{i-1}p_ip_{i+1}$. Similarly, in case of $P = 24$ five points are considered to choose the gray value of a point.

4.1 Building Reference Models

It has been observed that most of the historical documents composed with three basic regions: background (white or black), text and image. As an initial target, for separating text region from background and image, three reference models are considered from each type of region: i) background model ii) text model and iii) image model. The background models are built from the white and black portions of the images, image models are built from the image portions while text models are learnt from the text portions of various collection of complex document images. For each of the three reference models using LBP, normalized values of each bin in LBP-hisotgram are used as the model parameters. The comparison of such a set of three reference models (image, text and background) are shown in the Figure 3. On the other hand, for each of the three reference models using variance, mean (μ) and standard deviation (σ) are the model parameters.

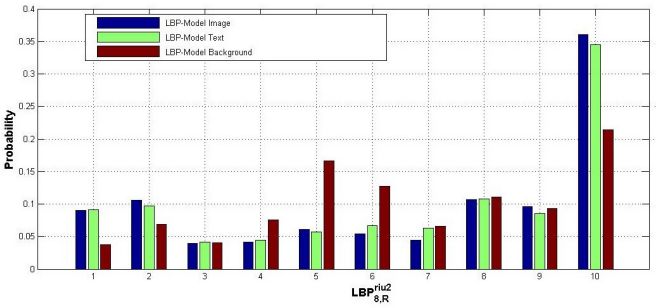


Fig. 3. Comparison of three reference LBP models (Image, Text and Background)

4.2 Measuring Model Response

Let us assume that $S_{w_{(i,j)}}$ is the sample by means of normalized LBP-histogram estimated from a certain window image $(w_{(i,j)})$. The model response for the sample S is defined as $L(S, M) = \sum_{b=1}^{B} S_b \log M_b$. It is also be noted that a block (a sample S) is assigned to the class of the model M that gives the maximum response. Similarly, the response of variance model is calculated from the normal distribution associated with the corresponding model parameters say mean (μ) and variance (σ^2).

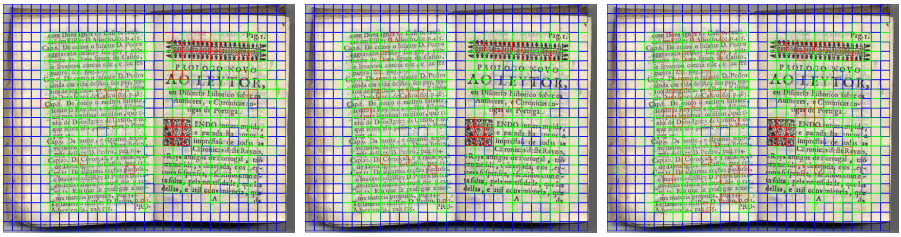
4.3 Block Re-labeling with Hierarchical Multiresolution Analysis

Initially, all the blocks are assigned to a certain label say, either it is background or text or image based on the model response. Let us assume that if a block is classified as "image", it is marked with number 0, if it is "text" then is marked

with number 1 and for "background" is marked with number 2. In block re-labeling process we re-assign the blocks further on the basis of the following criteria: i) If a block and its 8-neighbor blocks are in same label, then keep the label as it is. ii) If a block and its 8-neighbor blocks are in different labels but all the 8-neighbor blocks are in same label, then reassign the block with the neighbor's block label. iii) If a block and its 8-neighbor blocks are in different labels but all the 8-neighbor blocks are also not in same label, then all the neighbor's blocks are merged with its central block and resize it into previous block size, then extract the feature again from that block and examine the model response. If it is same as previous one, then the block label keep as it is, otherwise majority voting scheme is applied with the neighboring labels and change the block label accordingly.

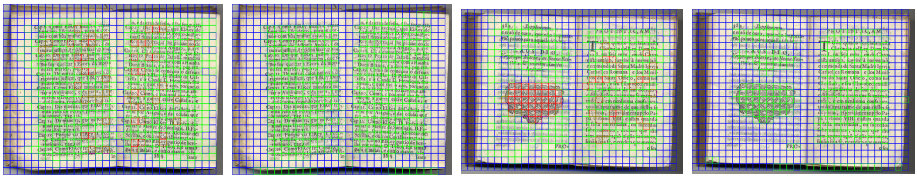
5 Experiments

First, we divide each document page logically into grids of a fixed window size. The size of the window is chosen based on the resolution of the image and it is calculated by the equation $w = \lfloor 3\lambda\sqrt{d} \rfloor$, where d is the resolution of the image measures in dpi and λ is a constant term. From our experiment, it is found that the localization scheme works well when $\lambda = 1$. Separate experiments have been performed using LBP and variance model. Some initial segmentation results are shown in the Figures 4 and 5. In the figures, the background, text and image regions are marked with blue, green and red windows respectively. Though from Figure 4 it is shown that the model $LBP_{8,R}$ gives slightly better result than the other two models $LBP^{ri}_{8,R}$ and $LBP^{riu2}_{8,R}$, but in many cases it is found that all the three LBP models produce almost similar results. However, from the experiments, we observe that the variance models are very effective to separate text and image regions from background but sometimes it fails to separate image region from the text (see in Figure 5(d)). On the other hand, when image regions (specific to document corpus, like as historical documents) have well defined number of texture patterns, LBP is very useful to separate image region from the text, even though both the text and image regions have similar variance. Such cases are shown in Figure 5. It is a fact that the performance of segmentation depends on the size of the window. So fixed-size window approach sometimes cannot handle documents that have text lines with capital letters or large fonts (such as heading, title) and it is likely to be classified as background region. To overcome this situation, we re-assign the label further with multi-resolution analysis (described in section 4.3). In addition, the re-labeling process helps to remove many erroneous labeling as well. Few such interesting results of the final segmentation (after re-assigning the label) are shown in the Figure 6(a) and Figure 6(b). We have evaluated our algorithm on 20 images ($2 * 20 = 40$ pages) from the archive of Portuguese historical documents [1] and overall performance of LBP and variance models are given in Table 1.



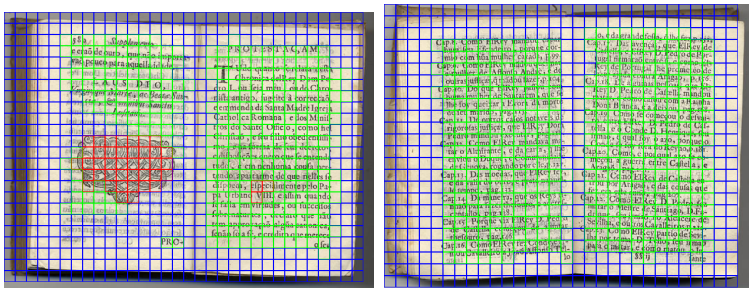
(a) (b) (c)

Fig. 4. Page segmentation of Portuguese document with LBP models. The outputs (a), (b) and (c) are produced by the three LBP operators $LBP_{8,R}$, $LBP^{ri}_{8,R}$ and $LBP^{riu2}_{8,R}$ respectively. The regions image, text and background are marked with red, green and blue colors respectively.



(a) (b) (c) (d)

Fig. 5. Page segmentation of Portuguese document with LBP models. The outputs (a), (b) and (c) are produced by the three LBP operators $LBP^{riu2}_{8,R}$ and var respectively. The regions image, text and background are marked with red, green and blue colors respectively.



(a) (b)

Fig. 6. Page segmentation of Portuguese document with (a) LBP models and (b) Variance models. Blue color indicates the background region and green color represent the text region.

6 Conclusion and Future Scope

In this paper, we explain a text localization scheme for historical documents with local binary pattern and variance model. We also introduce reference models concept in the intermediate stage to explore more difficulties for this kind of

Table 1. Text localization results on the archive of historical documents

No. of pages	No. of text zones	LBP		Variance	
		Recall (%)	Precision (%)	Recall (%)	Precision (%)
40	74	97.78	88.60	76.72	93.61

problems. As an initial experiment, three reference models background, text and image models are used from each category to identify the text region from background and image. The novelty of this framework is that we can add many more reference models depending upon the complexity of the document's archive. It is not necessary that only single text model can be used, multiple text models can be created from different classes of text textures. However, the response values from different models can be used to generate another level of features for further classification. In this stage, many sophisticated classification techniques such as SVM, MLP can be applied to separate text regions from the documents more accurately.

References

1. Biblioteca Nacional De Portugal, `http://purl.pt/index/geral/PT/index.html`
2. Baird, H.: Digital libraries and document image analysis. In: Proc. of the 7th ICDAR, pp. 2–14 (2003)
3. Etemad, K., Doermann, D., Chellappa, R.: Multiscale segmentation of unstructured document pages using soft decision integration. Trans. on. IEEE 19(1), 92–96 (1997)
4. Fletcher, L.A., Kasturi, R.: A robust algorithm for text string separation from mixed text/graphics images. Trans. on. IEEE 10(6), 910–918 (1988)
5. Gorman, L.O.: The document spectrum for page layout analysis. Trans. on. IEEE 15(11), 1162–1173 (1993)
6. Jain, A., Bhattacharjee, S.: Text segmentation using gabor filters for automatic document processing. Machine Vision Appl. 5, 169–184 (1992)
7. Kim, K.I., Jung, K., Kim, J.H.: Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. Trans. on. IEEE 25(12), 1631–1639 (2003)
8. Nagy, G., Seth, S.C., Stoddard, S.D.: Document analysis with an expert system. In: Pattern Recognition in Practice II, pp. 149–155. Elsevier Science, New York (1986)
9. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classi cation with local binary patterns. Trans. on. IEEE 24(7), 971–987 (2002)
10. Wong, K.Y., Casey, R.G., Wahl, F.M.: Document analysis system. IBM J. Res. Development 6, 456–642 (1982)
11. Wu, V., Manmatha, R., Riseman, E.M.: Textfinder: An automatic system to detect and recognize text in images. Trans. on. IEEE 21(11), 1224–1228 (1999)
12. Zheng, Y., Li, H., Doermann, D.: Machine printed text and identification in noisy document images. Trans. on. IEEE 26(3), 337–353 (2004)

Identification of *Devnagari* and *Roman* Scripts from Multi-script Handwritten Documents

Pawan Kumar Singh*, Ram Sarkar, Nibaran Das, Subhadip Basu, and MitaNasipuri

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
{raamsarkar,nibaran,bsubhadip,mitanasipuri}@gmail.com,
pawansingh.ju@gmail.com

Abstract. In a multilingual country like India it is a common scenario that a handwritten text document may contain more than one script. This causes practical difficulty in digitizing such a document, because the language type of the text should be pre-determined, before feeding it into a suitable Optical Character Recognition (OCR) system. In this paper, an intelligent feature based technique is reported, which automatically identifies the scripts of handwritten words from a document page, written in *Devnagari* script mixed with *Roman* script. The word-level script identification is performed by applying Multi layer Perceptron (MLP) based classifier with 39 distinctive features. The technique is tested on 100 handwritten document pages containing both *Devnagari* and *Roman* script words and 99.54% of words are identified with their true class.

Keywords: Script identification, Multi-script handwritten pages, Optical Character Recognition, Convex-hull feature, MLP classifier.

1 Introduction

India is a multi-lingual country with 25 official languages derived from 12 different scripts. A document page like railway reservation forms, question papers, language translation books and money-order forms etc. may contain words in more than one script/language. Each script has its own characteristics which is very different from other scripts. It is perhaps impossible to design a single recognizer which can identify a variety of scripts/languages. Thus, it is necessary to identify the language/script of the documents before feeding it to the corresponding OCR system. Script identification aims to extract information presented in digital documents namely articles, newspapers, magazines and e-books.

Besides, being the national language, *Hindi* is the most popular language in Indian sub-continent. *Devnagari* script is used to write *Hindi*, *Nepali*, *Marathi* and *Sindhi* languages. *English*, written in *Roman* script, has proven to be the binding language due to the diversity of languages and scripts in India and other countries in the world. For this reason, in the Indian sub-continent, people frequently use *English* words along with their native language in their writing. So, a script identification system to identify the script of the documents is of pressing need. This has motivated us to

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 509–514, 2013.

design an intelligent script identification technique which can successfully serve the immediate required purpose.

From the literature, it is revealed that in the context of Indian languages, some amount of research work on script/language identification has been reported [1-8]. Despite these research contributions, relatively few works [3, 5-8] are found on word-level script identification from handwritten bilingual script containing *Devnagari* script mixed with *Roman* script. In one of the previous works, R. Sarkar et al. [8] discussed about word-level script identification from *Bangla* and *Devnagari* handwritten script mixed with *Roman* script by using MLP classifier. In this paper, we propose a word-level script identification scheme based on 39 distinct features to identify scripts from the documents containing both *Devnagari* and *Roman* scripts by using MLP classifier.

2 Design of Feature Set

In the present work, 39 distinct features are designed for the identification of *Devnagari* and *Roman* script words using MLP classifier. These feature values, used with suitable normalization, are described as follows:

2.1 Matra/Shirorekha Feature

If the longest horizontal run of black pixels on the rows of a word is computed then such run length for *Devnagari* script will be much higher than that of *Roman* script. This is because characters in a word are generally connected by headline/Shirorekha/Matra in *Devnagari*. This information [14] is used to separate the words written in *Roman* script from *Devnagari* script.

2.2 Segmentation Based Feature

Here, two distinct features are considered *viz.*, number of Matra pixels and number of segmentation-point pixels [14].

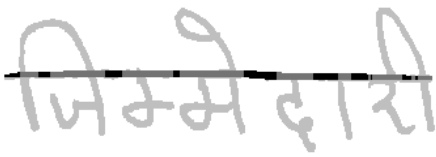


Fig. 1. Illustration of Matra pixels (light grey) and segmentation-point pixels (dark grey) are shown on a sample *Devnagari* word. [9]

Since *Roman* script words do not have Matra region, these pixels counts are significantly lower in case of *Roman* script in comparison to the words written in *Devnagari* script. Matra and segmentation-point pixels (see Fig. 1), are obtained by applying the character segmentation algorithm developed in [9].

2.3 Foreground-Background Transition Feature

The changeover of foreground and background pixels (transition point count) [13] is considered as feature value along 5 row positions in a particular word. For this purpose, the top row and bottom row of the word image are selected as R_1 and R_5 respectively. The row with maximum horizontalness is selected as R_2 as described in an earlier work on character segmentation[13].Then, the foreground-background changeover has been counted on 5 different row positions (see Fig. 2) such as R_2, $R_4=(R_2+R_5)/4$, $R_3=(R_2+R_4)/3$, $R_{12}=(R_1+R_2)/2$ and $R_{13}=(R_{12}+R_2)/2$ [8].

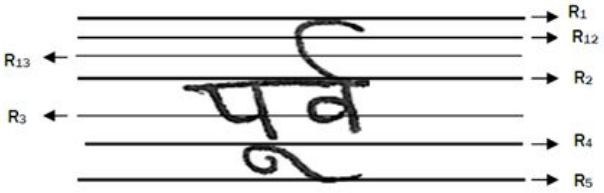


Fig. 2. Selection of specific rows for estimation of foreground-background feature values. [8]

2.4 Convex-Hull Based Feature

The Convex hull of a set of points S in n dimensions is the intersection of all convex sets containing S. For $\boldsymbol{N}$ points $\boldsymbol{P_1, P_2, \ldots, P_N}$, the convex hull C is then given by the expression

$$\mathbf{C} = \left\{ \sum_{\mathbf{j=1}}^{\mathbf{N}} \boldsymbol{\lambda_j p_j} : \boldsymbol{\lambda_j \geq 0 \text{ for all } j \text{ and } \sum_{j=1}^{N} \lambda_j = 1} \right\} \quad (\mathbf{1})$$

In the current work, we have used Graham scan algorithm [13] for computing the convex hull of binary pattern of each word image. From the Green's theorem [14], it can be shown that the area A of convex hull is given by

$$\mathrm{A} = \frac{1}{2} \sum_{\mathrm{i=1}}^{\mathrm{L}} (\mathrm{x_i\, y_{i+1} - x_{i+1} y_i}) \quad (2)$$

where, L is the number of order vertices, (x_i, y_i) coordinates of the order vertices forming polygon.

Also, the centroid (C_x, C_y) of the convex hull can be expressed as:

$$C_x = \frac{1}{6A} \sum_{i=1}^{L} (x_i + x_{i+1})\,(x_i y_{i+1} - x_{i+1} y_i) \quad (3)$$

$$C_y = \frac{1}{6A}\sum_{i=1}^{L}(y_i + y_{i+1})\,(x_i y_{i+1} - x_{i+1} y_i) \tag{4}$$

Now, the deficit of convexity can be defined as the set of pixels inside the convex hull of any object pattern which does not belong to the said object. There may be two types of convex deficiencies *viz*., regions totally enclosed by the object, called *lakes* and regions lying between the convex hull perimeter of the object, called *bays* [10]. It has been seen from the convex hull structure of words that the number of *lakes* found in case of *Devnagari* words is more than that of the *Roman* words. On the other hand, in case of *Roman* script words, number of *bays* is more than that in *Devnagari* script words. In the current work, we have extracted 31 basic topological features (like *bays* attributes, *lakes* attributes, etc.) from the convex hull of handwritten word images. For extracting local information, each such word pattern is further divided into four sub-images based on the *centroid* of its convex hull. After that, new convex hulls have been constructed for each such sub-image.

In the present work, d_{cp} are calculated as the column and row wise distances of data pixels from the convex hull boundary taken from the top, bottom, right and left boundaries of any image. Now, the seven topological features are calculated as maximum d_{cp}, total number of rows having $d_{cp} > 0$, average d_{cp}, mean row coordinate having $d_{cp} > 0$, total number of rows having $d_{cp} = 0$, number of visible bays from left to right direction and, convex hull perimeter feature (total number of convex hull pixels having $d_{cp} = 0$ from four sides). From the top, bottom, right and left boundaries of the image (7x4=28) such features are calculated. Finally, the remaining three features *viz*., total numbers of *lakes*, *bays* and convex hull perimeter pixels having $d_{cp} > 0$ are taken under consideration. This makes the total feature count as 39.

3 Experimental Results

For the experimental purpose, a database of 100 handwritten pages comprising of 13941 words are considered. Handwritten document pages have been collected from different types of sources, viz., class notes of students of different age-groups, and the document pages, written by different persons, on request. The document pages written under supervision were collected from various persons with varied textual contents of the books containing both *Hindi* and *English* vocabularies. The documents are digitized by a HP scanner at 300 dpi. The digitized images are in gray tone and we have used a histogram-based thresholding approach to convert them into two-tone images. Both the *Hindi* and *English* word images from the document pages are then manually cropped for the evaluation of the present technique.

We have used 3-fold cross validation scheme for evaluating the script recognition algorithm. The detail of the experimental setup is shown in Table 1. For developing a trained network for each of the MLP based classifiers, several runs of BP learning algorithm with learning rate (η) = 0.7, momentum term (α) = 0.7 and adjustment factor =0.9 are executed for different number of neurons in its hidden layer. The accuracies of 3 different runs of script identification scheme are 99.27%, 99.54% and 99.40% respectively.

Only few words are misclassified during test. The main reasons for misclassification of *Devnagari* words are due to discontinuities in Matra and poor quality of documents due to presence of noise (see Fig. 3(a)). Some of the *Roman* script words are also misclassified (see Fig. 3(b)). The possible reason may be due to existence of Matra like component in the word, extracted feature values are identical to the word written in *Devnagari* script. In addition, the presence of some small component in the upper part for *Devnagari* and *Roman* scripts, sometimes they are misclassified among each other.

Table 1. Detail results of the present script identification technique

Scripts	Set #1		Set #2		Set #3	
	Number of words trained	Number of words tested	Number of words trained	Number of words tested	Number of words trained	Number of words tested
Devnagari	7500	1992	7508	2002	7430	2014
Roman	3410	1031	3412	1026	3520	1011
TOTAL	10910	3023	10920	3028	10950	3025
Success rate for test case (%)	99.27		**99.54**		99.40	

(a) *Devnagari* script (b) *Roman* script

Fig. 3. Sample word images where present technique fails to identify the actual script

4 Conclusion

Script identification has been the forerunner for developing a multi-script OCR system. In this paper, a scheme for word-level handwritten script identification technique is presented. Document pages, applied in the present work, are written in *Devnagari* script mixed with *Roman* script. An intelligent set of 39 features with MLP classifier is applied to identify the scripts of the words. With 3-fold cross validation of 13941 words, the present technique classifies 99.54% words with their true classes. In addition to this, the scheme also works well for single character words. The script identification technique, in future, can also be applied to other scripts like *Bangla*, *Gurmukhi*, etc. mixed with *Roman* script because of the similar script structure. Also, this scheme will be effective and efficient for different writing styles of handwriting which ensures generality of the present technique. In future, we aim to improve the accuracy of the technique by minimizing the script dependency of the features. In addition to this, we are also planning to design a general script identification module for the development of multi-script OCR system.

References

1. Pal, U., Choudhuri, B.B.: Script Line Separation From Indian Multi-Script Documents. In: Proc. of 5th International Conference on Document Analysis and Recognition, pp. 406–409. IEEE Comput. Soc. Press (1999)
2. BasvarajPatil, S., Subba Reddy, N.V.: Character script class identification system using probabilistic neural network for multi-script multi lingual document processing. In: Proc. of National Conference on Document Analysis and Recognition, Karnataka, pp. 1–8 (2001)
3. Pal, U., Choudhuri, B.B.: Automatic Separation of Words in Multi Lingual multi Script Indian Documents. In: Proc. of 4th International Conference on Document Analysis and Recognition, pp. 576–579 (1997)
4. Chanda, S., Pal, U.: English, Devnagari and Urdu Text Identification. In: Proc. of International Conference on Document Analysis and Recognition, pp. 538–545 (2005)
5. Pal, U., Sinha, S., Choudhuri, B.B.: Word-wise script identification from a document containing English, Devanagari and Telugu text. In: Proc. of 2nd National Conference on Document Analysis and Recognition, Karnataka, India, pp. 213–220 (2003)
6. Padma, M.C., Nagabhushan, P.: Identification and separation of text words of Kannada, Hindi and English languages through discriminating features. In: Proc. of 2nd National Conference on Document Analysis and Recognition, Karnataka, pp. 252–260 (2003)
7. Spitz, A.L.: Determination of the script and language content of document images. Proc. of IEEE Tran. on Pattern Analysis and Machine Intelligence 19, 234–245 (1997)
8. Sarkar, R., Das, N., Basu, S., Kundu, M., Nasipuri, M., Basu, D.K.: Word-Level Script Identification from Bangla And Devanagri Handwritten Texts mixed with Roman script. Journal of Computing 2(2) (2010)
9. Basu, S., Sarkar, R., Das, N., Kundu, M., Nasipuri, M., Basu, D.K.: A Fuzzy Technique for Segmentation of Handwritten Bangla Word Images. In: Proc. of International Conference on Computing: Theory and Applications (ICCTA), pp. 427–433 (2007)
10. Das, N., Pramanik, S., Basu, S., Saha, P.K., Sarkar, R., Kundu, M., Nasipuri, M.: Recognition of handwritten Bangla basic characters and digits using convex hull based feature set. In: Proc. of International Conference on Artificial Intelligence and Pattern Recognition, AIPR (2009)
11. Vysniauskaite, L., et al.: A Priori Filtration Of Points For Finding Convex Hull, Tede, vol. XII(4), pp. 341–346 (2006)
12. Lady, E.L.: (February 14, 2000), `http://www.math.hawaii.edu/~lee/calculus/green.pdf`
13. Sarkar, R., Das, N., Basu, S., Kundu, M., Nasipuri, M., Basu, D.K.: A two-stage approach for Segmentation of Handwritten Bangla word Images. In: Proc. of International Conference on Frontiers in Handwritten Recognition (ICFHR), Canada, pp. 403–408 (2008)
14. Sarkar, R., Malakar, S., Das, N., Basu, S., Kundu, M., Nasipuri, M.: Word Extraction and Character Segmentation from Text Lines of Unconstrained Handwritten Bangla Document Images. Journal of Intelligent Systems 20(3), 227–260 (2011)

Extraction of Doodles and Drawings from Manuscripts

Chandranath Adak[1] and Bidyut B. Chaudhuri[2]

[1] Dept. of CSE, University of Kalyani, West Bengal-741235, India
adak32@gmail.com
[2] CVPR Unit, Indian Statistical Institute, Kolkata-700108
bbc@isical.ac.in

Abstract. In this paper we propose an approach to separate the non-texts from texts of a manuscript. The non-texts are mainly in the form of doodles and drawings of some exceptional thinkers and writers. These have enormous historical values due to study on those writers' subconscious as well as productive mind. We also propose a computational approach to recover the struck-out texts to reduce human effort. The proposed technique has a preprocessing stage, which removes noise using median filter and segments object region using fuzzy c-means clustering. Now connected component analysis finds the major portions of non-texts, and window examination eliminates the partially attached texts. The struck-out texts are extracted by eliminating straight lines, measuring degree of continuity, using some morphological operations.

Keywords: Connected Component, Document Image Analysis, Doodle Separation, Fuzzy C-Means Clustering, Manuscript Processing.

1 Introduction

In the field of *document image analysis* [1], the separation of texts and non-texts has gained interest since 1980. It is important, so that they can be sent to different systems/engines for processing. The texts are fed to OCR and non-texts are sent to graphics processing system. The texts may be printed, handwritten and mixture of both (hybrid). The existing methods [2-10] deal with different logos, diagrams, maps, engineering drawings and photographic images.

Here we separate the doodles and drawings from ancient manuscripts. For our experiment, we choose manuscripts of Leonardo da Vinci, Gustave Flaubert, Lewis Carroll, Rabindranath Tagore and Samuel Beckett. In the manuscript of Vinci, there are some engineering models and human figures. Some crossed lines and rectangular boxes are present in Flaubert's manuscript. Carroll's manuscript contains human figures and Beckett's manuscript is full of funny characters. In Tagore's manuscript, we find doodles of various shapes (e.g. real and imaginary animals, trees, human models, phantoms etc.).

The challenge of our work is that most of the doodles and drawings are touching the texts, and sometimes the doodles are formed with struck-out lines

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 515–520, 2013.

(ink-strokes) of irregular patterns. We approach to recover the texts behind those struck-out lines.

In our preprocessing stage, we use median filter to remove noise and fuzzy c-means clustering (FCM) to segment the image into object regions. We apply connected component (CC) analysis to separate non-touching doodles, examine windows to detach touching non-texts (doodles) and extract struck-out texts using straight line elimination, degree of continuity measurement and mathematical morphological operations.

2 Proposed Method

The proposed method consists of following steps:

1. The manuscript is scanned in RGB color (I_{rgb}, fig.1.a) and converted into gray-scale image I_{gray} (fig.1.b).
2. Noise is removed from I_{gray} using 3-by-3 neighborhood *median filtering* to get the image I_{nf} (fig.1.c).
3. I_{nf} is segmented (fig.1.d) by *FCM clustering* to obtain the ink-strokes and background pixels separated. The background pixels are converted to zero value (black), while the foreground ink-strokes are converted into one (white). Let the resulting image is I_{bin}.
   Steps 1-3 are basically preprocessing stage.
4. We generate all the connected components of white pixels from I_{bin}. The doodles are usually dense, well-connected, large sized components. We extract the larger connected components (doodles) by a threshold T (fig.1.e). Now non-touching doodles (those do not touch any text) and texts are separated (fig.1.f-g).
5. For text touching with doodles, we note the basic features of text that it is elongated with curvy, thin, smaller lines with variation of degree of continuity. On I_{bin} we take a 5×5 window, the pixel values of this window is zero/one due to binarization. If all pixels of a row/column/diagonal are one, then it is more likely that the window is part of doodle (fig.2.b-c).
   Steps 4,5 are repeated interactively for a satisfactory outcome.
6. Next we identify and delete the struck-out lines, which is done as follows:
   (a) Morphological *thin* and *shrink* [11] operations are used on the remaining image.
   (b) We eliminate the horizontal, vertical, diagonal straight lines by examining 5×5 window and checking *degree of continuity*.
   (c) After eliminating those straight lines, small pieces of lines may remain. We compute CC on the remaining image. If the number of pixels in a CC is less than a threshold, then that CC is converted into background.
   (d) In the remaining image, there is the skeleton of text that was struck-out. We find the edge of this skeleton using *Sobel* operator. We create dilated image I_d by morphological *dilate* [11] operation. On I_d we return the gray value of the original image I_{gray} (fig.4.b,e).
   Steps 4,5,6 are repeated for a suitable separation of texts and non-texts.
7. We combine these all non-touching, touching and struck-out texts to obtain the total text portion. The remaining portions are considered as non-texts.

3 Experimental Result and Discussion

To assess the stability and correctness of the proposed method, the results are obtained from different ancient manuscripts.

In *fig.1*, we handle the trivial non-touching case, where the non-texts (doodles) are not connected with the texts.

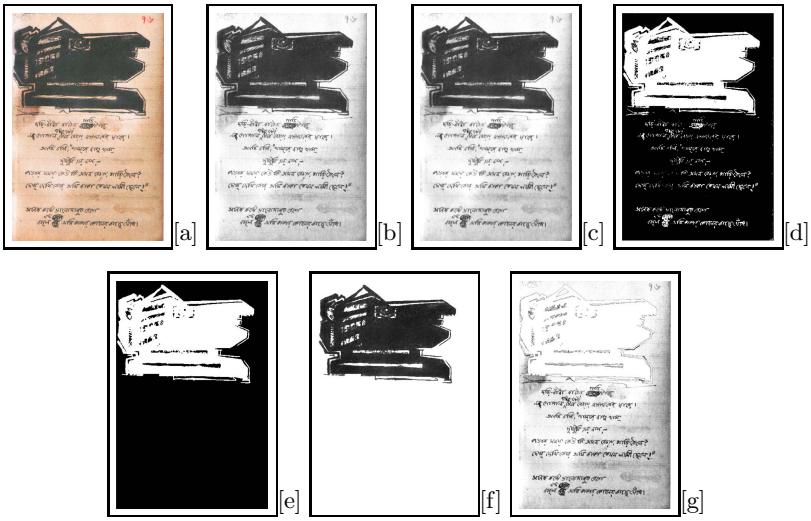


Fig. 1. *Non-touching case:* Manuscript of Tagore's poem ("Nutan Srota"), *Parisesh*, 19 August, 1927: (a) original, (b) gray, (c) noise free, (d) segmented (binary), (e) CC (large), (f) gray value of CC (doodle), (g) texts in gray

In *fig.2*, we separate the touching texts and non-texts by the window examination. *Fig.2.c* shows the doodle in binary after eliminating the major portion of touching text.

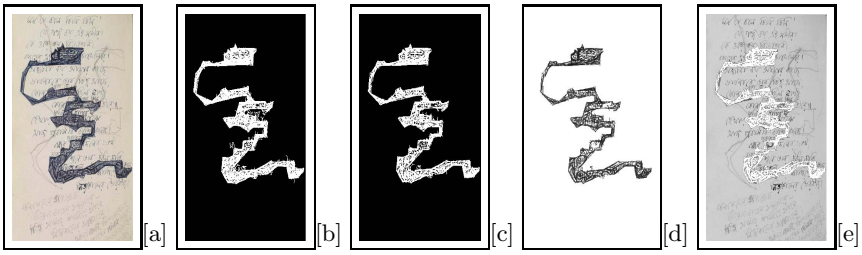


Fig. 2. *Touching case:* Manuscript of Tagore's song, *Geetabitan*, 1929: (a) original, (b) larger CC: before window examination, (c) CC: after window examination, (d) CC in gray (doodle), (e) texts in gray

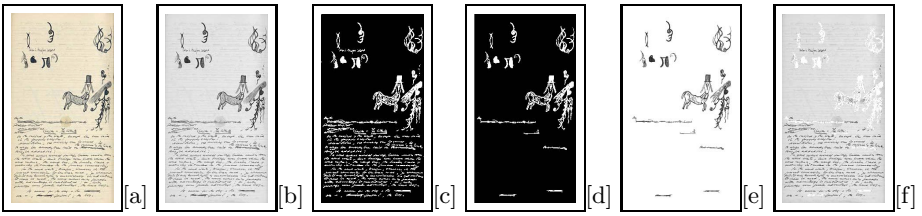


Fig. 3. Beckett's manuscript, *Watt I*, 1941: (a) original, (b) noise free gray, (c) segmented (binary), (d) CC (large), (e) CC in gray (doodle), (f) texts in gray

Our proposed method also works for small doodles on Beckett's manuscript. It is shown in *fig.3*. Some bold struck-out ink-strokes are marked as non-texts.

We approach to extract the struck-out texts on Tagore's manuscript shown in *fig.4.a,c*. Total recovery is not possible by our proposed method. After extracting the struck-out text, we combine this with the remaining texts.

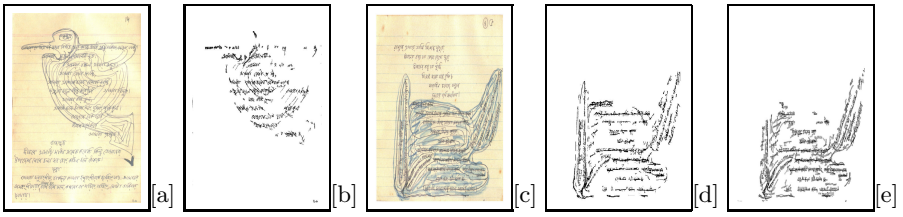


Fig. 4. *Struck-out case:* Tagore's manuscript: (a) original (*Tasher Desh*, 1933), (b) struck-out texts in gray scale, (c) original (*Nabajatok*, 4 May 1939), (d) struck-out texts in binary, (e) struck-out texts (with changing threshold) in gray

In *fig.5*, we deal with Vinci's manuscript and separate texts and drawings.

Fig.6 shows complex (where human cannot read the struck-out texts properly without a great effort) case analysis of our proposed method.

We take total 115 manuscripts of different sizes, out of which there are 24 non-touching, 63 touching, 23 struck-out and 5 complex cases.

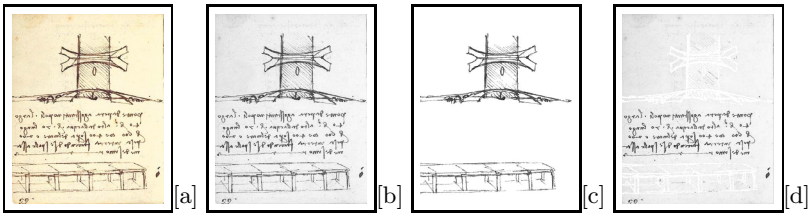


Fig. 5. Vinci's manuscript, *Golden Horn Bridge design*, 1502: (a) original, (b) noise free gray, (c) drawings in gray, (d) texts in gray

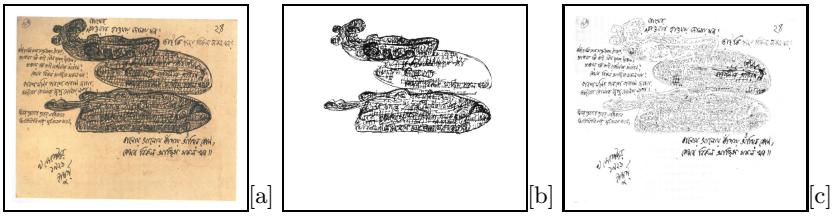


Fig. 6. *Complex case:* Manuscript of Tagore's song ("Chokher chaoar haoay dolay mon"), *Geetabitan*, 9 Sep. 1926: (a) original, (b) doodle in gray, (c) texts in gray

For performance analysis of our proposed method, we calculate precision (P), recall (R) and F-measure (F) considering their standard definition.

In table 1, we show the average P, R and F for different cases. To calculate the average, we use the arithmetic mean of P, R and F for all images under each case.

The *true positive (TP), true negative (TN), false positive (FP)* and *false negative (FN)* are defined as follows:
TP: # pixel in the actual doodle portion in an image (correctly classified),
TN: # pixel in the actual text portion (correctly classified),
FP: # pixel in text incorrectly labeled as doodle (unexpected result),
FN: # pixel in doodle incorrectly marked as text (missing result).

Table 1. Performance analysis of doodles and drawings extraction

Case Study	Non-touching	Touching	Struck-out	Complex
P %	99.80	98.71	60.31	20.78
R %	99.28	99.15	59.52	84.92
F %	99.54	98.93	59.91	33.39

The non-touching and touching cases show 99.54% and 98.93% F-measure respectively, the struck-out case shows 59.91% F-measure, but for complex case it is only 33.39% , we have found less number of struck-out and complex cases, mostly they are touching.

To the best of our knowledge, a preliminary work on such separation was done by Chaudhuri et al. [12] on simple doodles of Tagore's manuscript and they obtained 92.17% accuracy (90.14% F-measure). We did not get any other reference/work to do further comparative study.

4 Conclusion

A text and non-text (doodle) separator from ancient manuscript has been presented in this paper. The proposed algorithm works well for handwritten/printed

texts and it is independent of scripts. This algorithm can be extended to some other applications in document image analysis, such as image restoration from ancient documents, wills and testaments, newspapers, magazines, articles etc. Though it works for touching and non-touching cases, it does not work well for complex cases, and more modification is required to explore struck-out texts. Our system is semi-automatic, so our next venture will be to make the system automatic and more accurate.

References

1. Nagy, G.: Twenty Years of Document Image Analysis in PAMI. IEEE Trans. on PAMI 22(1), 38–62 (2000)
2. Luo, H., Agam, G., Dinstein, I.: Directional Mathematical Morphology Approach for Line Thinning and Extraction of Character Strings from Maps and Line Drawings. In: Proc. ICDAR 1995, pp. 257–260 (1995)
3. Kasturi, R., Bow, S.T., El-Masri, W., Shah, J., Gattiker, J.R., Mokate, U.B.: A System for Interpretation of Line Drawings. IEEE Trans. on PAMI 12(10), 978–992 (1990)
4. Dori, D., Liu, W.: Vector-Based Segmentation of Text Connected to Graphics in Engineering Drawings. In: Perner, P., Rosenfeld, A., Wang, P. (eds.) SSPR 1996. LNCS, vol. 1121, pp. 322–331. Springer, Heidelberg (1996)
5. Lu, Z.: Detection of Text Regions from Digital Engineering Drawings. IEEE Transactions on PAMI 20(4), 431–439 (1998)
6. Fletcher, L.A., Kasturi, R.: A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images. IEEE Tran. on PAMI 10(6), 910–918 (1988)
7. He, S., Abe, N.: A Clustering-Based Approach to the Separation of Text Strings from Mixed Text/Graphics Documents. In: IEEE Proc. of ICPR 1996, pp. 706–710 (1996)
8. Adak, C.: Unsupervised Text Extraction from G-Maps. In: Proc. Int. Conf. on Human Computer Interactions (ICHCI 2013), India (August 2013)
9. Roy, P.P., Lladós, J., Pal, U.: Text/Graphics Separation in Color Maps. In: Proc. Int. Conf. on Computing: Theory and Applications (ICCTA 2007) (2007)
10. Garg, R., Hassan, E., Chaudhury, S., Gopal, M.: A CRF Based Scheme for Overlapping Multi-Colored Text Graphics Separation. In: Proc. ICDAR 2011, pp. 1215–1219 (2011)
11. MATLAB R2012a (7.14.0.739), MathWorks Inc., `http://www.mathworks.com`
12. Chaudhuri, B.B., Borah, S., Saraf, A., Goyal, A., Kumari, A.: Separation of Text from Non-Text Doodles of Poet Rabindranath Tagore's Manuscripts. In: Proc. Nat. Conf. on Comp. and Comm. Systems (NCCCS 2012), pp. 1–5 (November 2012)

Text Segmentation from Land Map Images

Samit Biswas and Amit Kumar Das

CST Department, BESU, Shibpur, Howrah, India
samitbiet@yahoo.com, amit@cs.becs.ac.in

Abstract. Text segmentation from scanned map images is a challenging task as the texts in maps have wide variation in font, skew, spatial spread, touching components and a myriad background. We present a segmentation process by separating map components to three layers based on their areas. Further analysis, on spatial spread and node density from component graph leads to segmentation with encouraging performance.

1 Introduction and Related Work

Maps are ideal sources for getting geographical, historical, social and political information of a particular period. These pieces of information are mostly embedded in the maps as texts. However, the task of text extraction in maps is far more complex than other types of documents. The presence of multiple fonts, text touching with other line/line like structures representing other components, complex background and orientation of the texts have made it particularly challenging. Due to the complex nature of the problem, literature on text extraction from map images is not as rich as text segmentation from documents of other types. Related major works are summarised next which would be a cue to the backdrop of the approaches taken so far.

Tofani and Kasturi [1] extracted texts from color map images. At first the Authors' separate individual color overlays or alternatively separated each feature overlays (e.g., annotation, contour line, physical boundaries land use area shading etc.) based on their color. Then black/pink pixel layer of the map image and text extracted based on area thresholding was done. Li et al. [2] separated texts and line arts from USGS topographic maps by analysing the connected components. The maps were consists of street lines and labels. The constrained considered for component separation were defined such as space width between two characters(1.4 times their average width), characters of a text resides on a line, street casings (supposed that all street casings were connected in a map, not broken) were of 0.1 mm wide and separated 0.5 mm. Cao et al. [3] detected characters from complex documents (i.e., characters overlaps lines) using the differences of the length of line segments in character and line. Authors manually cut out the neighboring area of overlapping text with lines. The obtained piece of images were used to separate text and lines. Roy et al. [4,5] separated the maps into different layers according to color features and then connected component features, different size thresholds were used for this initial layering and

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 521–529, 2013.

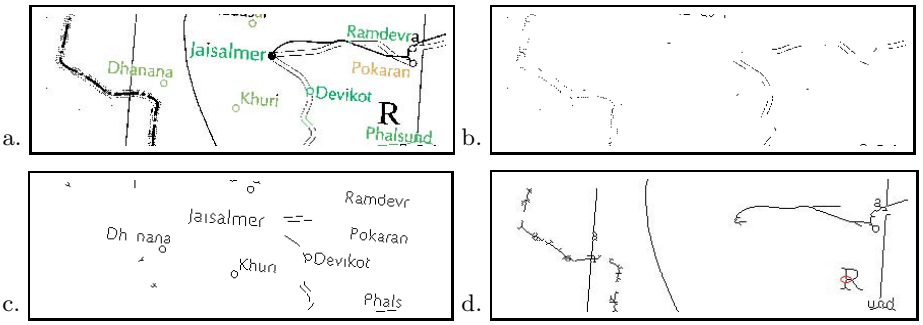


Fig. 1. a) Map image, colored texts are extracted using our approach; **b)** First layer; **c)** Middle layer; and **d)** Last layer; A junction point is shown encircled with red color

then skeleton information were used to identify text characters. Hough Transform was used to detect long straight lines. Curved lines separated by analyzing the length of major axis of decomposed components. Pezeshk and Tutwiler [6,7] used Multi Angled Parallelism (MAP) as a tool to separate text and graphics layer in maps. Authors are introduced a line representation method and a set of directional morphological operations to extract the text layer from intersecting linear features in scanned map images. In [8,9,10] authors extracted text from map images based on Radon transform based projection profile, CC analysis and graph based modeling respectively. These worked well for map images which contain *Bangla* text.

2 Proposed Approach

Map components are placed in three layers (first, middle and last) based on the size of the components. This layering is similar to the work reported in [1,4,5]. However, the novelty here is in the unique treatment of the components in middle and last layer based on spatial distribution of components (in the middle layer) and forming a graph by selecting nodes (pixels) on the components in the last layer. Scanned gray image, I_{gr} is first binarized (I_{bw}) and then thinned (I_{th}). I_{th} is analysed using the areas of the connected components (CC) to form multiple layers. Following subsections explain layering, analysis of each layer and combining the missing components.

2.1 Layer Formation Based on the Area of Each Component

The CCs are divided into three layers based on its area only. *First* layer contains most of the small components (such as dots, dashes, etc.). Map components which belongs to texts (words) reside within a moderately bigger range and are grouped in the *middle* layer. The *last* layer contains most of the big components; like boundary lines, rivers, portion of texts touched with other elements etc.

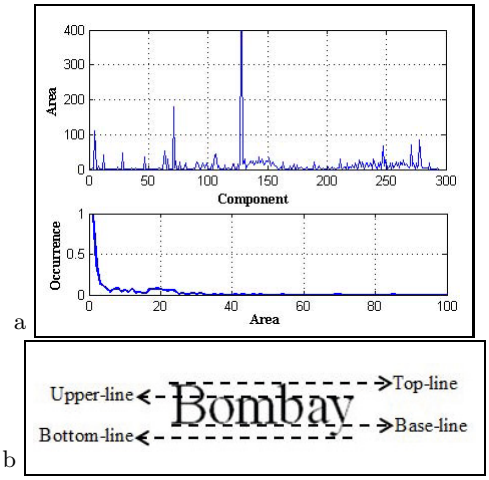


Fig. 2. a) Upper half shows Area of each CC; Lower half shows the histogram of the area of CCs; b) Spatial distribution of characters of text

Now, compute the area of each CC from I_{th}. Upper half of Fig.-2a shows the area correspond to each component of Fig.-1a. The histogram based on the area of those CCs is formed (see lower half of Fig.-2a). The histogram has one highest peak and the threshold (T_{h1}) for extracting components for the *first* layer lies at valley on the right side of it. CCs whose areas are less than T_{h1}, belong to first layer. CCs lying in the *middle* layer is grouped by another threshold, T_{h2} computed by taking the average area of all components excluding the components from the first layer. Middle layer components have got area in the range of T_{h1} to T_{h2} and components whose areas are crossing the threshold t_{h2} have been put to the last layer. For example see Fig.-1) which fairly supports these basis for forming three different layers.

The CCs of *first* layer are removed because most of these are noises. The *middle* layer is analysed because the components are isolated and have to be grouped together forming words or part of a word. As noted earlier few text portions may be grouped to the last layer as they are touching a bigger component. After grouping the connected components for each portion of text contained in *middle* layer, there is a need to get back the texts which may have been, inadvertently, put to the *last* layer during layer formation. This requirement can be observed from Fig-1c and 1d where portions of the texts have gone to the last layer as they are touching big components (lines in this case). Similarly, small lines, circles and similar components have been placed, wrongly, in the middle layer.

2.2 Analysis of Middle Layer and Distance Profile Generation

The characters belonging to a word are grouped based on the Distance Profile (DP). Before elaborating the DP we draw the attention of some characteristics of the text. The characters forming a word (most of the cases) are isolated

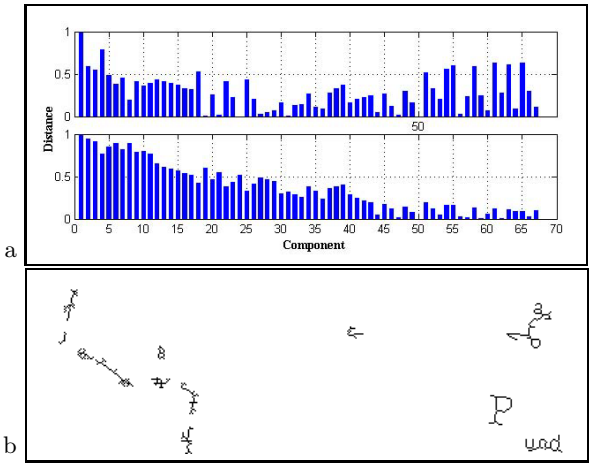


Fig. 3. a) Horizontal and vertical DP for i^{th} component, (here i=50); b) Modified last layer, Fig.-1d; text and text like components are retained

components. The main body of a character is largely confined within the base-line to top-line while the extended portion (outliers) coming out of the body is confined either between the upper-line and top-line or between the base-line and bottom-line (see Fig.-2b). Thus the top most points and bottom most points of each letter of a text reside on near vicinity of the top and base lines, respectively.

The analysis of the middle layer CCs is aimed at grouping discrete characters to a single unit (words in this case). This grouping is done by generating a DPs for each component which is a measure of the horizontal and vertical distances of the other components from the concerned component. At first for a component we find the topmost and bottom most points and drawing a straight line through these two points. Now we draw a straight line perpendicular to the first line and passing through the bottom most point. The second line is considered to be the base line for that particular component. Next, we draw perpendiculars to this base line from the bottom most points of all other components. The perpendicular distances thus obtained reflect the vertical distances of all other components with respect to the chosen component. Similarly, the distance of each perpendicular line projected on the base line reveal the horizontal distances of all other components with respect to the first component. Fig.-3a shows DP for a component; upper half shows the vertical distances while the lower half shows the horizontal distances. We generate the DPs for each and every component.

A group is formed around a component by taking all other components whose horizontal and vertical distances are less than or equal to lengths of major, m_1 and minor, m_2 axis of the concerned component, respectively. Finally, a regrouping among these groups have been done to find the chain connected set of components which is likely to be a word(s) or part of it.

2.3 Analysis of *Last* Layer

The last layer consists ideally the big components; mostly lines. Some of these lines may have touched a character and dragged the character to the last layer as an integral part of its own. The objective of analysing the last layer is to get back those characters from last to middle layer. First compute the nodes that can be formed from each component. A pixel is considered to be a node if it has one or more than two neighbours.The terminal points as well as the junction points are considered as nodes according to this logic. Note that for a thinned component most of the pixels are a part of a continuous run and are neither the terminal pixel or the junction pixel and they will have exactly two neighbours; which may be termed as left and right or top and bottom. We keep the node pixels of all the components removing all non-node pixels. In the next step form a weighted graph by connecting each node to all other nodes.

Edge weight (W_{ij}) between two nodes (v_i and v_j) is computed as per Equation-1 where number of neighbors at node v is $N(v)$ and $d(v_i, v_j)$ is the Euclidean distance between two nodes, v_i and v_j. The node density will be more in junction points (e.g., where a character is touching a line) and the edges will get more weights for short edges connecting closely spaced node cluster (junction points). This graph is partitioned using the edge weight as threshold to remove edges with low weight. Note that the edge weights will exponentially reduce and the threshold value may be set easily observing the knee point [9].

$$W_{ij} = \begin{cases} 1 & \text{if} \quad v_i = v_j \\ \frac{1}{d(v_i,v_j)}[1 - \frac{1}{N(v_i)+N(v_j)}] & \text{if} \quad N(v_i) \text{ and } N(v_j) \neq 1 \\ 0 & \text{Otherwise} \end{cases} \tag{1}$$

Consider a partition and project all the nodes back to the component(s) in last layer. If there is a continuity between any two nodes we keep that part of the thinned component of the last layer. This way the characters (with some extraneous part) which have been put to last layer may be segmented from the usual line like components of the last layer (compare Fig.-3b with Fig.-1d).

2.4 Combining the Missing Components

The characters have grouped to form words in middle layer. We also have isolated the characters connected to line(s) and were pushed to this layer truncating some of the words of the middle layer. These are to be combined now and for that purpose we have generated search strip from the extreme points of the components in the middle layer as elaborated next.

The extrema points (top-left, top-right, bottom-left, bottom-right, left-top, left-bottom, right-top and right-bottom) of all components from each subset (group of components) in the middle layer are computed first. Now we connect the top-left and bottom-left most extreme points of the leftmost component of a group; the perpendicular lines passing through the top-left and bottom-left

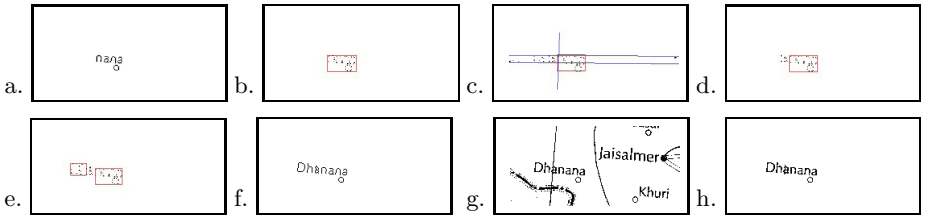


Fig. 4. a) A group of CCs considered from middle layer, b) Extreme points of the considered portion ('nana') of a word; c) Search strip; d) Extrema points grouped within the strip, e)After regrouping, marker-point set for a text; f) Associated set of connections among marker points (used as marker image, I_c); g) Original binarized input map image (used as mask image, I_{bw}), h)Reconstructed text, $R_C(I_c)$

Algorithm 1. Text Reconstruction from initial connections, I_c among marker-points of a set

Require: Marker, (I_c) and Mask, (I_{bw}).
Ensure: Reconstructed Text, $R_C(I_c)$.
1: Initialize h_1 to be the marker image, I_c
2: Create the structuring element: B
3: $Area_1$ = Total number of marker pixels in h_1; $C_1 = 0$
4: **Repeat** the following:
$h_{k+1} = (h_k \bigoplus B) \bigcap I_{th}$; $Area_{k+1}$=Total number of marker pixels in h_{k+1}
$C_{k+1} = Area_{k+1} - Area_k$
Until $C_{k+1} \geq C_k$
5: $R_C(I_c) = h_{k+1}$
6: **return** $R_C(I_c)$;

point on the connecting line specifies a band and has been extended across the image. Exactly the same thing is done on the rightmost component of a group. The two bands would roughly overlap and would make our search strip. All the extrema points of all components from middle and last layer within this search strip are stored together. The actions described are shown in Fig.-4a to c.

Next, we form two graphs using; i) all the extrema points of the components of the group and ii) extrema points under the search strip including the points of the group (that may not exclusively confined within the strip) and put edge weight using the Euclidean distance between the nodes. From the first graph we divide the edge weights into ten equally spaced bins and select the bin as threshold which has got maximum number of edge weights. Now from the second graph delete edges and corresponding nodes which are more than the selected threshold. This way we get a set of groups possibly enhanced with extreme points coming from other groups of middle and last layers. We combined these enhanced groups by merging them with common members (extreme points). Finally, a conditional morphological reconstruction (see Algorithm-1), using the extrema points of the groups, is done to form the text masks.

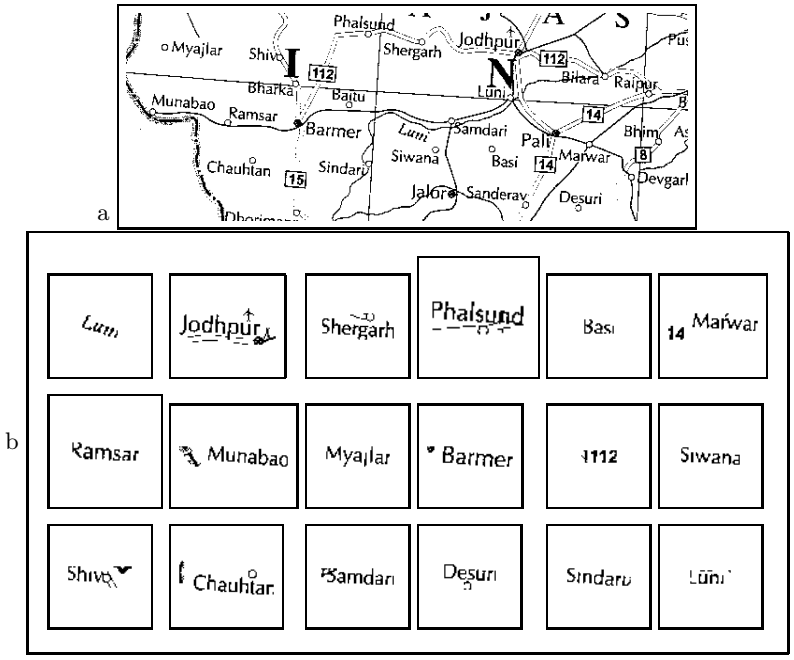


Fig. 5. a) A portion of a map image; b) Text images segmented (few are shown)

Table 1. Performance of the proposed method

No. of Map Images	True positive(%)	MDP(%)	IDP(%)	False Positive(%)
117	86.05	10.28	3.46	11.28

3 Dataset Details and Experimental Result

We have created our own dataset (BESUMAPS) with a variety of map images collected from sources like *websites*, *map books*[1] and *land surveyors*. The collected paper maps are scanned at 200 dpi in TIFF format. Each of the test map image consists of texts with myriad background consists of various intensity values, orientations, overlapping objects, intersected lines etc. At present the population in the BESUMAPS is 157 out of which 117 are in English and rest in different Indian languages including *Bangla*. The texts (place names etc.) have been manually extracted for each map and a ground-truth is made available. Fig.-5a shows a portion of a map which has been used as input to the proposed approach, while Fig.-5b shows the output.

Overall performance is shown in Table-1 for maps in English from BESUMAPS. *True Positive* (TP), *False Positive* (FP), *Incomplete Detection Percentage* (IDP) and *Mixed Detection Percentage* (MDP) are used for performance measure. TDP

[1] Oxford School Atlas, Oxford University Press.

indicates the extractions where texts have only been identified while MDP indicates texts with extraneous components. IDP gives an account of the texts (words) where we have missed a portion of the text (a character or part of it) and FP counts the extracted component which contain no text. Considering TP and MDP together the results are certainly encouraging. FP is rather high due to the presence of many small components which when grouped exhibits text-like characteristics.

4 Conclusion

In maps the major problem is that the lines do not merely touch the character; they pass through them and make them an integral part of the texts. This is an inherent problem and may be difficult to separate with the available algorithms generally used to solve the so called "touching character" problem. We are currently working towards a practical solution to dissociate the strongly bonded lines (or similar components) from the characters for the improved performance of our extractor. In the context of non-availability of well known map image databases and the small size of our database we are consolidating our efforts on creating a comprehensive database along with the ground truth. This would enable us to compare our proposed algorithm with the other algorithms already reported in the literature.

References

1. Tofani, P., Kasturi, R.: Segmentation of text from color map images. In: Proceedings of the Fourteenth International Conference on Pattern Recognition, vol. 1, pp. 945–947 (August 1998)
2. Li, L., Nagy, G., Samal, A., Seth, S., Xu, Y.: Cooperative text and line-art extraction from a topographic map. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition, ICDAR 1999, pp. 467–470 (September 1999)
3. Cao, R., Tan, C.L.: Text/graphics separation in maps. In: Proceedings of the 4th International Workshop on Graphics Recognition Algorithms and Applications, pp. 245–254 (2001)
4. Roy, P.P., Vazquez, E., Lladós, J., Baldrich, R., Pal, U.: A system to segment text and symbols from color maps. In: Liu, W., Lladós, J., Ogier, J.-M. (eds.) GREC 2007. LNCS, vol. 5046, pp. 245–256. Springer, Heidelberg (2008)
5. Roy, P., Llados, J., Pal, U.: Text/graphics separation in color maps. In: International Conference on Computing: Theory and Applications, ICCTA 2007, pp. 545–551 (2007)
6. Pezeshk, A., Tutwiler, R.L.: Improved multi angled parallelism for separation of text from intersecting linear features in scanned topographic maps. In: ICASSP, pp. 1078–1081 (2010)
7. Pezeshk, A., Tutwiler, R.L.: Automatic feature extraction and text recognition from scanned topographic maps. IEEE T. Geoscience and Remote Sensing 49(12), 5047–5063 (2011)

8. Biswas, S., Das, A.K.: Text segmentation from scanned land map images using radon transform based projection profile. In: 2011 International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 413–418 (October 2011)
9. Biswas, S., Das, A.K.: Fuzzy graph modeling for text segmentation from land map images. In: Proceedings of the Eighth Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2012, pp. 75:1–75:7 (2012)
10. Biswas, S., Das, A.K.: Text extraction from scanned land map images. In: IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision, Dhaka, pp. 231–236 (May 2012)

Rough-Fuzzy Clustering and M-Band Wavelet Packet for Text-Graphics Segmentation

Pradipta Maji, Shaswati Roy, and Malay K. Kundu

Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
{pmaji,shaswatiroy_t,malay}@isical.ac.in

Abstract. This paper presents a segmentation method, integrating judiciously the merits of rough-fuzzy computing and multiresolution image analysis technique, for documents having both text and graphics regions. It assumes that the text and non-text regions of a given document are considered to have different textural properties. The M-band wavelet packet is used to extract the scale-space features, which is able to zoom it onto narrow band high frequency components of a signal. A scale-space feature vector is thus derived, taken at different scales for each pixel in an image. Finally, the rough-fuzzy-possibilistic c-means algorithm is used to address the uncertainty problem of document segmentation. The performance of the proposed technique, along with a comparison with related approaches, is demonstrated on a set of real life document images.

1 Introduction

With the advances in information technology, automated processing of documents has become an imperative need. The documents in digitized form require a large amount of storage space, after being compressed using advanced techniques. Text-graphics segmentation partitions a document image into distinct regions corresponding to the text and non-text parts facilitating efficient searching and storage of the text parts in documents.

Many techniques have been proposed to segment the document image into text and non-text regions in the literature [1]. Recently, wavelet techniques have become powerful tools in this domain. Li and Gray [2] have used features based on distribution characteristics of wavelet coefficients in high frequency bands to segment document images into four classes, namely, background, photograph, text, and graph. Kundu and Acharyya [3] proposed a scheme for text-graphics segmentation based on wavelet scale-space features followed by k-means clustering. Lee et al. [4] used an algorithm based on local energy estimation in wavelet packet domain and k-means clustering.

In this paper, a text-graphics segmentation method is proposed, which integrates the principles of rough-fuzzy computing and multiresolution image analysis technique. This approach is based on the assumption that the text portion of the document image is comprised of one texture class and the non-text part of the other. The M-band wavelet packet (MWP) is used to extract the scale-space features, which offers a richer range of possibilities for document image

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 530–538, 2013.

and is able to zoom it onto narrow band high frequency components of a signal. It yields a large number of subbands which are required for good-quality segmentation. Subsequently features are computed by using nonlinear energy estimation followed by a smoothing filter. However, one of the main problems in document image segmentation analysis is uncertainty. It includes incompleteness and vagueness in class definitions. To address this issue, an unsupervised clustering algorithm, termed as rough-fuzzy-possibilistic c-means (RFPCM), is used to segment the feature vectors. This approach needs not to assume any a priori information regarding the font size, scanning resolution and type of layout.

2 Rough-Fuzzy-Possibilistic C-Means Algorithm

Let $X = \{x_1, \cdots, x_j, \cdots, x_n\}$ be the set of n objects and $V = \{v_1, \cdots, v_i, \cdots, v_c\}$ be the set of c centroids and β_i be the ith cluster, where $x_j \in \mathbb{R}^m$ and $v_i \in \mathbb{R}^m$. Each cluster in the RFPCM algorithm [5] is represented by three parameters, namely, a cluster centroid, a crisp lower approximation and a fuzzy boundary. The centroid is calculated based on the weighting average of the crisp lower approximation $\underline{A}(\beta_i)$ and fuzzy boundary $B(\beta_i)$ as follows:

$$v_i = w \times \mathcal{C}_1 + \tilde{w} \times \mathcal{D}_1; \text{ where } \mathcal{C}_1 = \frac{1}{|\underline{A}(\beta_i)|} \sum_{x_j \in \underline{A}(\beta_i)} x_j, \tag{1}$$

$$\mathcal{D}_1 = \frac{1}{n_i} \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{m\acute{}_1} + b(\nu_{ij})^{m\acute{}_2}\} x_j, n_i = \sum_{x_j \in B(\beta_i)} \{a(\mu_{ij})^{m\acute{}_1} + b(\nu_{ij})^{m\acute{}_2}\}.$$

Here $1 \leq m\acute{}_1 < \infty$ and $1 \leq m\acute{}_2 < \infty$ are the fuzzifiers, $\mu_{ij} \in [0, 1]$ and $\nu_{ij} \in [0, 1]$ are the probabilistic and possibilistic membership functions, respectively, of the object x_j to the cluster β_i. The parameters w and $\tilde{w}(= 1 - w)$ correspond to the relative importance of lower approximation and boundary region, respectively. The constants a and $b(= 1 - a)$ define the relative importance of probabilistic and possibilistic memberships, respectively. The probabilistic and possibilistic membership values of an object x_j are calculated as

$$\mu_{ij} = \left(\sum_{k=1}^{c} \left(\frac{d_{ij}^2}{d_{kj}^2} \right)^{\frac{1}{m\acute{}_1 - 1}} \right)^{-1}; \text{ where } d_{ij}^2 = ||x_j - v_i||^2, \tag{2}$$

$$\text{and } \nu_{ij} = \frac{1}{1 + E}; \text{ where} E = \left\{ \frac{b||x_j - v_i||^2}{\eta_i} \right\}^{\frac{1}{m\acute{}_2 - 1}}. \tag{3}$$

The scale parameter η_i represents zone of influence of cluster β_i. In the RFPCM, the membership values of objects in lower approximation are $\mu_{ij} = \nu_{ij} = 1$, while those in boundary region are the same as fuzzy-possibilistic c-means. The main steps of the RFPCM algorithm are as follows

1. Assign initial centroids v_i, $i = 1, 2, \cdots, c$ and the fuzzifiers $\acute{m}_1$ and $\acute{m}_2$.
2. Compute μ_{ij} and ν_{ij} by (2) and (3), and finally, u_{ij} for c clusters and n objects where $u_{ij} = \{a\mu_{ij} + b\nu_{ij}\}$.
3. Calculate threshold δ, which determines the class labels of all objects as
$$\delta = \frac{1}{n}\sum_{j=1}^{n}(u_{ij} - u_{kj})$$
where n is the total number of objects, u_{ij} and u_{kj} are the highest and second highest memberships of x_j.
4. If $(u_{ij} - u_{kj}) \leq \delta$, then $x_j \in \overline{A}(\beta_i)$ and $x_j \in \overline{A}(\beta_k)$, where $\overline{A}(\beta_i)$ represents the upper approximation of cluster β_i. Furthermore, x_j is not part of any lower approximation region.
5. Otherwise, $x_j \in \underline{A}(\beta_i)$. In addition, by properties of rough sets, $x_j \in \overline{A}(\beta_i)$.
6. Modify μ_{ij} and ν_{ij} as 1 for the objects in lower approximations, while those in boundary regions are remain unchanged.
7. Compute new centroid as per (1).
8. Repeat steps 2 to 7, until no more new assignments can be made.

3 Feature Extraction

This section presents the feature extraction methodology that includes multi-channel filtering using the MWP with adaptive basis selection and subsequently local energy estimation and smoothing.

3.1 *M*-Band Wavelet Packet

In dyadic wavelet (2W) [6], scaling ϕ and wavelet functions ψ are defined as

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k); \text{ and } \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{4}$$

for all $j, k \in \mathbb{Z}$. Here k determines the position along x-axis; j determines function's width, that is, how broad or narrow it is along x-axis.

The M-band orthonormal wavelet bases are constructed as a direct generalization of the 2W. The 2W decomposes a signal into frequency subbands that have the same bandwidth on a logarithmic scale, whereas M-band wavelet (MW), in addition, focuses on narrow band high frequency components of a signal, thereby simultaneously having a logarithmic and a linear decomposition of frequency channels. Let $\phi(t)$ be the scaling function satisfying

$$\phi(t) = \sum_k h_\phi(k)\sqrt{M}\phi(Mt - k). \tag{5}$$

Additionally, the $M - 1$ wavelets can be expressed as

$$\psi_l(t) = \sum_k h_\psi^l(k)\sqrt{M}\psi(Mt - k); l = 1, \cdots, M - 1, \tag{6}$$

where $h_\phi(n)$ and $h_\psi^l(n)$ are scaling and wavelet function coefficients, respectively. Scaling and translating the functions, $\phi(t)$ and $\psi_l(t)$, the $\phi_{j,k}(t)$ and $\psi_{l,j,k}(t)$ are obtained, respectively, as

$$\phi_{j,k}(t) = M^{j/2}\phi(M^j t - k), \tag{7}$$

$$\psi_{l,j,k}(t) = M^{j/2}\psi_l(M^j t - k); l = 1, \cdots, M-1. \tag{8}$$

For any given function $f(t) \in L^2(\mathbb{R})$, it can be shown that

$$f(t) = \sum_{k\in\mathbb{Z}} \langle f(t)\phi_{j,k}(t)\rangle \, \phi_{j,k}(t) + \sum_{l=1}^{M-1} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \langle f(t)\psi_{l,j,k}(t)\rangle \, \psi_{l,j,k}(t) \tag{9}$$

where $\mathbb{Z}$ represents the set of integers, $l = 1, \cdots, M-1$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}$ and $< ., . >$ represents the inner product.

An MW decomposes an image into M^2 subbands. Since in the MWP, at each decomposition level every subband is further decomposed, each of these M^2 subbands gives rise to another M^2 number of bases. So if the decomposition depth is p, then an MWP decomposition results in M^{2p} number of subbands, and this large number of bases are required for good quality segmentation.

3.2 Filtering and Adaptive Basis Selection

In the filtering stage, an eight-tap four-band orthogonal wavelet [7], as shown in Table 1, is used to decompose the document images into $M \times M$ channels without downsampling. The filter length is increased with the increasing level of decomposition. The filters are expanded by inserting appropriate number of zeros between taps of filters, thereby satisfying the quadrature mirror filter condition.

Table 1. Filter Coefficients for Eight-Tap Four-Band Wavelet

# of Taps (n)	$\phi(n)$	$\psi_1(n)$	$\psi_2(n)$	$\psi_3(n)$
0	-0.067371764	-0.094195111	-0.094195111	-0.067371764
1	0.094195111	0.067371764	-0.067371764	-0.094195111
2	0.40580489	0.56737176	0.56737176	0.40580489
3	0.56737176	0.40580489	-0.40580489	-0.56737176
4	0.56737176	-0.40580489	-0.40580489	0.56737176
5	0.40580489	-0.56737176	0.56737176	-0.40580489
6	0.094195111	-0.067371764	-0.067371764	0.094195111
7	-0.067371764	0.094195111	-0.094195111	0.067371764

In order to find out a suitable basis without going for a full decomposition, an adaptive decomposition algorithm using a maximum entropy or information content criterion extracted from each of the subbands is used [8]. After first level decomposition of the image into $M \times M$ channels, energy for each subband is evaluated. Among these subbands, those for which energy values exceed a threshold value (ϵ_1) of the parent band energy, are considered and decomposed further. A subband at second decomposition level is further decomposed if its energy value is more than another threshold value (ϵ_2) of the total energy of all the subbands at current scale. Hence, the number of subbands can be generated in this scheme in the range of 16 to 4096. Empirically it is seen that $\epsilon_1 = 0.01$ and $\epsilon_2 = 0.10$ are good choice for the images considered in the experiment.

3.3 Local Energy Estimation and Smoothing

After the selection of significant bases, a local estimator, which constitutes a nonlinear operator followed by a smoothing filter, is applied to each subbands [3]. It gives high energy value for the regions in each subbands where frequency components are strong, otherwise low energy value is obtained where it is weak.

In this feature-extraction scheme, standard deviation is used as the nonlinear operator, calculated over small overlapping windows around each pixel. The local energy $E_b(x, y)$ around the (x, y)th pixel of bth subband is given as

$$E_b(x, y) = \sqrt{\frac{1}{R} \sum_{m=1}^{w} \sum_{n=1}^{w} \left| F_b(m, n)^2 - \overline{F}_b(x, y)^2 \right|} \tag{10}$$

where w is the window size and $R = w \times w$. $\overline{F}_b(x, y)$ is the mean around the (x, y)th pixel and $F_b(m, n)$ is the filtered image. Gaussian low-pass filter used as the smoothing filter is of the form

$$H_G(u, v) = \frac{1}{2\pi\sqrt{\sigma}} e^{-\frac{1}{2\sigma^2}(u^2+v^2)} \tag{11}$$

where σ determines the passband width of the averaging filter. Formally, the feature image $Feat_b(x, y)$ corresponding to subband image $F_b(x, y)$ is given by:

$$Feat_b(x, y) = \frac{1}{G^2} \sum_{(m,n) \in G_{xy}} \Gamma(F_b(m, n)) H_G(x - m, y - n) \tag{12}$$

where $\Gamma(.)$ gives the energy measure and G_{xy} is a $G \times G$ window centered at pixel with coordinates (x, y). It is found that an averaging window size of 9×9 to be appropriate in most of the segmentation experiment.

3.4 Choice of Energy Window Size

A nonlinear energy estimator is used in order to discriminate texture pairs as the unprocessed wavelet coefficients do not convey enough information for efficient representation of texture cues. In the present work, the energy window size is decided based on the measure of the edge density of the image as follows:

$$Den_e = \frac{no.\ of\ edge\ pixels}{total\ no.\ of\ pixels\ in\ image} \tag{13}$$

It gives the measure of overall image busyness. Sobel edge detector is used here to extract the edges. Den_e has a dynamic range of values between $[0, 1]$. For highly active image, Den_e is close to 1, so a smaller window for nonlinear operation is required. Moderately active images having the value of Den_e within $[0, 1]$ would require a moderate window size for good feature extraction. It has been found experimentally that the energy window size, for these different categories of image activities, ranges from 5×5 to 19×19.

4 Experimental Results

In this section, the performance of proposed text-graphics segmentation methodology is extensively compared with wavelet based different feature extraction techniques and several clustering algorithms. In the current study, features are extracted by the 2W decomposed upto third level, dyadic wavelet packet (2WP) decomposed upto third level, the MW as implemented in [3] and the proposed MWP. Daubechies 6 filter has been used for the feature extraction using the 2W and 2WP decomposition. The filter coefficients are shown in Table 2. The MW and MWP decompositions use the filter coefficients as depicted in Table 1. The clustering algorithms involve hard c-means (HCM), fuzzy c-means (FCM) [9], possibilistic c-means (PCM) [10], fuzzy-possibilistic c-means (FPCM), rough-fuzzy c-means (RFCM) [5], rough-possibilistic c-means (RPCM) [5] and RFPCM [5]. The values of parameters $\acute{m}_1 = \acute{m}_2 = 2.00$, $a = 0.50$ and $w = 0.95$.

Table 2. Filter Coefficients for Six-Tap Daubechies Wavelet

# of Taps (n)	$\phi(n)$	$\psi(n)$
0	0.3326705530	-0.0352262919
1	0.8068915093	-0.0854412739
2	0.4598775021	0.1350110200
3	-0.1350110200	0.4598775021
4	-0.0854412739	-0.8068915093
5	0.0352262919	0.3326705530

Some of the document images analyzed in the experiment are standard documents, others are scanned and taken online from parts of Anandabazar Patrika (www.anandabazar.com) and Times of India (timesofindia.indiatimes.com). Structured document images with nonoverlapping text and graphics regions are shown in Fig. 1(i) having size of 496×496, Fig 1(ii) of 256×256, Fig 1(iii) of 377×431 and Fig 1(iv) of 512×512.

Fig. 1(v) - Fig. 4 show the comparative analysis among text-graphics segmentation algorithms using different feature extraction methods to prove the efficacy of the proposed algorithm. From the results reported in Fig. 1(v) - Fig. 4, it is seen that there is a significant improvement in the segmentation results using the MWP compared to the classical 2W and 2WP, where $M = 2$. This may be explained by the significance of intermediate frequency bands obtained using the MWP decomposition in characterizing the textural features. Here, in these figures, the performance of different clustering techniques is also analyzed. It is found that rough set based clustering approaches are yielding good segmentation results over non-rough set based algorithms, irrespective of feature extraction techniques and images. Among all rough set based clustering methods, the RFPCM gives excellent results as far as text identification is concerned.

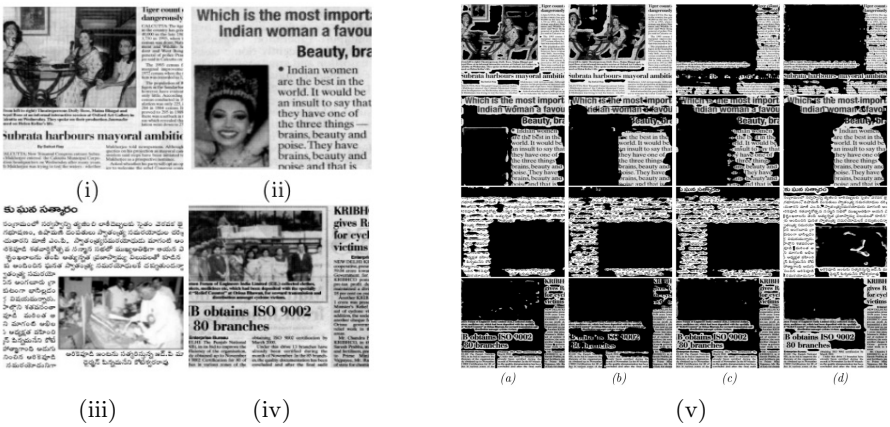


Fig. 1. (i)-(iv) Input data set for document image segmentation. (v) Text obtained using HCM: (a) 2W (b) 2WP (c) MW (d) MWP.

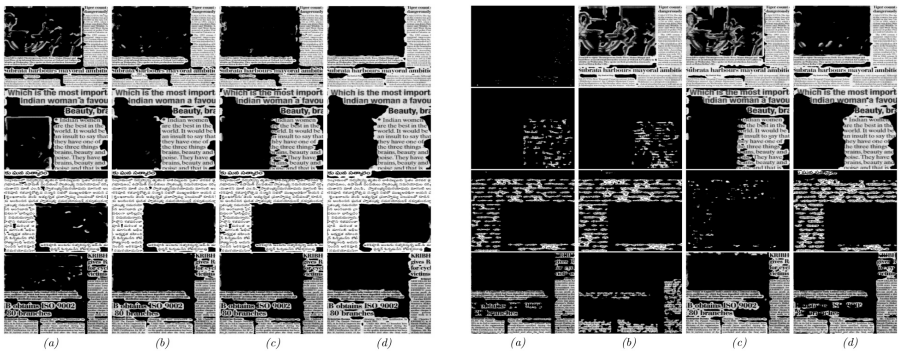


Fig. 2. Text obtained using FCM and PCM: (a) 2W (b) 2WP (c) MW (d) MWP

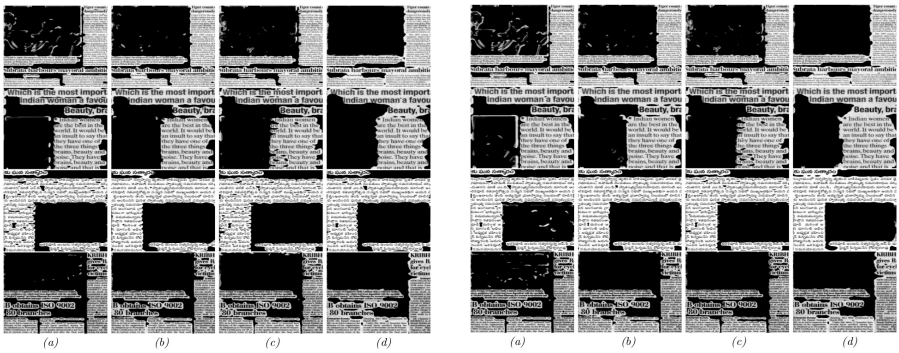


Fig. 3. Text obtained using FPCM and RFCM: (a) 2W (b) 2WP (c) MW (d) MWP

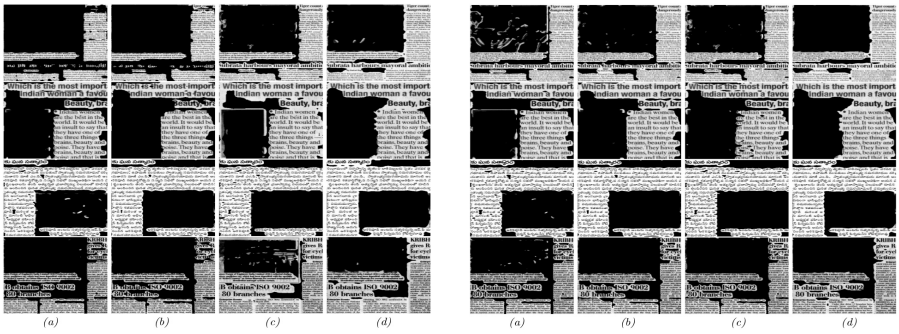


Fig. 4. Text obtained using RPCM and RFPCM: (a) 2W (b) 2WP (c) MW (d) MWP

5 Conclusion

In this paper, a new texture-based methodology is presented for segmenting the text part from the graphics part based on textural cues. The use of wavelet theory via MWP decomposition of images provides a multiscale multidirectional representation of the image and yields a huge number of frequency channels and hence facilitates an improved segmentation of the different class regions. The RFPCM is geared towards maximizing the utility of both rough sets and fuzzy sets with respect to uncertainty handling. An extensive comparative study with different feature extraction techniques and clustering approaches shows that the proposed methodology is indeed effective in characterizing document images in a better way.

References

1. Srihari, S.N.: Document Image Understanding. In: Proc. Fall Joint Computer Conference, pp. 87–96 (1986)
2. Li, J., Gray, R.M.: Context-Based Multiscale Classification of Document Images Using Wavelet Coefficient Distributions. IEEE Trans. Image Processing 9(9), 1604–1616 (2000)
3. Acharyya, M., Kundu, M.K.: Document Image Segmentation Using Wavelet Scale-Space Features. IEEE Trans. Circuits and Systems for Video Technology 12(12), 1117–1127 (2002)
4. Lee, G.B., Odoyo, W.O., Lee, J.H., Chung, Y., Cho, B.J.: Two Texture Segmentation of Document Image Using Wavelet Packet Analysis. In: Proc. 9th Intl. Conf. Advanced Communication Technology, vol. 1, pp. 395–398 (2007)
5. Maji, P., Pal, S.K.: Rough Set Based Generalized Fuzzy C-Means Algorithm and Quantitative Indices. IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics 37(6), 1529–1540 (2007)
6. Mallat, S.G.: Multifrequency Channel Decompositions of Images and Wavelet Models. IEEE Trans. Acoustics, Speech and Signal Processing 37(12), 2091–2110 (1989)

7. Alkin, O., Caglar, H.: Design of Efficient M-Band Coders With Linear-Phase and Perfect-Reconstruction Properties. IEEE Trans. Signal Processing 43(7), 1579–1590 (1995)
8. Acharyya, M., Kundu, M.K.: Image Segmentation Using Wavelet Packet Frames and Neuro-Fuzzy Tools. Intl. Jrnl. Computational Cognition 5(4), 27–43 (2007)
9. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Kluwer Academic Publishers Norwell, MA (1981)
10. Krishnapuram, R., Keller, J.M.: A Possibilistic Approach to Clustering. IEEE Trans. Fuzzy Systems 1(2), 98–110 (1993)

A New Image Binarization Technique by Classifying Document Images

Soumik Datta*, Pawan Kumar Singh, Ram Sarkar, and MitaNasipuri

Department of Computer Science and Engineering, Jadavpur University, Kolkata, India
{soumik.datta86,pawansingh.ju,raamsarkar,
mitanasipuri}@gmail.com

Abstract. The present work proposes a binarization algorithm based on classification of document images. The method first classifies the images into two categories namely, simple and complex images. The global threshold value is used for binarizing the simple document images whereas complex document images are binarized by applying local threshold values. A background checking method is introduced in this method to detect the blocks which can be marked as purely background blocks. Finally, a post-processing mechanism has been applied to improve the quality of the binarized image.

Keywords: Binarization, Document image analysis, Optical Character Recognition, Global Thresholding, Local Thresholding, Simple Document Images, Complex Document Images.

1 Introduction

We are standing in the fifth generation of computer - the age of automation and artificial intelligences. During the last four decades, attention was paid in document analysis through computer. Document analysis involves handwriting recognition, writer identification, signature verification, etc. Most of these algorithms take binarized image as input to reduce the complexity and computational cost of the algorithms. Thus, binarization is the forerunner of many image processing techniques used in document image analysis. It is the process of converting 256 levels of grayscale information into two levels (black and white) image information. To binarize a grayscale image at first, threshold value(s) has to be determined. If a pixel value is less than the threshold value, then the pixel value of the corresponding output image is set as 1 (black) otherwise it is set as 0 (white).

Binarization is a challenging task [1-4] when noise is contaminated with the document image due to various reasons. In case of degraded documents, often the background and data pixels are misclassified if optimal threshold(s) has not been determined by the binarization process. It becomes more challenging when the following two cases occur: (i) the image contains text and graphic, (ii) the image contains text objects having a wide variety of gray-levels.

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 539–544, 2013.

2 Previous Work

In the literature, a lot of research work has been found on binarization algorithms. The algorithms can be broadly classified into three categories depending on the estimation of threshold values, namely: global thresholding algorithms, local thresholding algorithms and hybrid thresholding algorithms. In case of global thresholding approach, a single threshold value is determined to binarize the whole document. Image Binarization algorithms developed by Otsu [5], Kittler and Illngworth [6] belong to this category.

On the other hand, in case of local thresholding approach, the document image is divided into sub-images and threshold values are determined for each sub-image. Popular local threshold based binarization algorithms are Bernsen [7], Niblack [8], Sauvola [9], etc. Many recent research works are going on to combine the results of both global and local thresholding schemes to produce better results by removing the drawbacks of both types of algorithms. These algorithms are referred to as hybrid binarization algorithms [10-11].

3 Present Work

Textual documents often contain noise, shadow and other types of degradations. On the other hand, some document images are also found which are almost noise free. Therefore, the proposed binarization algorithm works into three modules (steps) to deal with both categories of the images:

- **Module 1:** Categorization of the document image, whether it is simple or complex.
- **Module 2:** Binarize the document image depending on its type.
- **Module 3:**Perform the post processing operations to improve the quality of the binarized images.

3.1 Categorization of Document Images

The proposed algorithm, first classifies the document images into two classes depending on the histogram analysis of the images namely— 1) Simple Document Images (*SDI*) and 2) Complex Document Images (*CDI*). The document images with very less or no noisy pixels are classified as *SDI*, whereas, the document images with a large amount of noisy pixels are classified as *CDI*.

Histogram of an image, which is a useful tool for numerous spatial domain image processing techniques, shows the frequency distribution of its pixels. To determine the type of the document, a technique proposed by S. H. Shaikh et al. [11] is used. At first, the number of peaks in the histogram of the input image is determined. A gray value 'P' will be determined as peak, if its frequency is greater than that of its two previous and two next gray values. In the next step, the average of all the peak values is determined which is called average peak. The peak, whose frequency is greater than the average peak value, is called sharp peak. If the image has at most two sharp peaks then the image is marked as *SDI* otherwise, it is *CDI*. Fig. 1 shows all types of peaks.

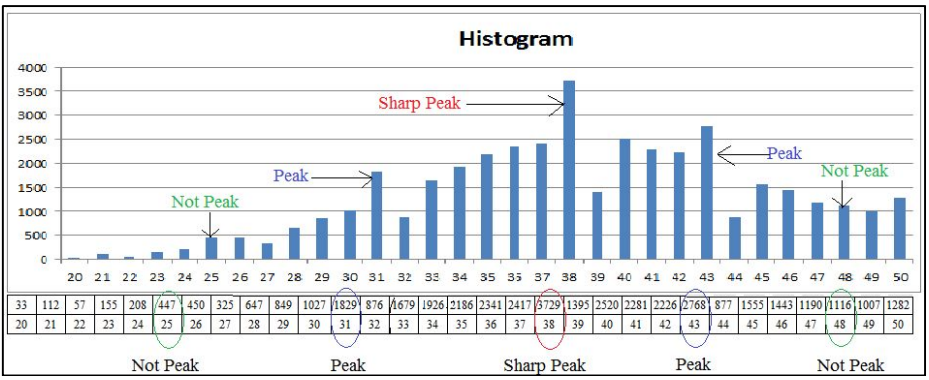


Fig. 1. Showing the peaks and sharp peaks in a part of an image histogram

3.2 Binarization of document images

- **Simple Image:** In case of *SDI*, the existence of at most two sharp peaks signifies (implies) that the intensities of the pixels possibly dispersed into two different categories. Among these, the first category implies the foreground whereas the second category implies the background. In this case, Otsu's methodology [5] is directly applied to binarize the document image.
- **Complex Image:** For *CDI*, the image has more than two sharp peaks which imply that the pixel intensities may be dispersed into more than two different categories. Thus, in this case Otsu's methodology may not be directly applied to binarize the document image. In the current work, the following technique is used to binarize such types of document images.

At first, the global threshold value as GTh for the whole document image is determined by using the following formula:

$$GTh = \frac{\text{Maximum PixelIntensity Value} + \text{Minimum PixelIntensity Value}}{2} \quad (1)$$

Next, the image is divided into blocks of NxN pixels. Then, maximum and minimum gray level values for each block are determined. After that *background blocks* are detected and merged. The process of detecting these blocks is known as background thinning. If all the pixels of a block belong to the background class then the block is referred to as *background block*. These blocks will not be considered for further processing. An image block is determined as a *background block* if the following two conditions are satisfied:

Condition 1: Maximum pixel intensity value must be greater than the global threshold for white background i.e., Maximum Gray Level Value > GTh

Condition2: The difference of the maximum and minimum pixel value must be less than or equal to B_t i.e., Maximum Gray Level Value – Minimum Gray Level Value ≤ B_t, where B_t is a constant.

The blocks other than *background blocks* are referred to as *data blocks* which are analyzed in the next step of the binarization process. Finally, the data blocks are binarized using Otsu's algorithm.

3.3 Post-processing

- **Post-processing of *SDI*:** To improve the quality of the binarized document image, a post-processing method has been introduced. For this, 8-connected neighbors of a particular data pixel in the original grayscale image are considered. If any of them is found to be a background pixel, then the grayscale value at that pixel position in the original image is checked. That pixel position is set as data pixel in the output binarized image if the said gray scale value is within 120% of the GTh. An extension of the threshold value is used to improve the image quality because some object pixels, become faded (whitish) due to degradation of the image, are misclassified as background pixels in the output image as the Otsu's Algorithm considers only the pixel intensity values and not the relation between the pixels.
- **Post Processing of *CDI*:** As *CDI* contain more noise, improving such images is more challenging. Another post processing operation, discussed below, is performed to remove those noisy pixels.

Step 1: Count the connected data pixel-clusters for each image block using flood fill algorithm.

Step 2: Repeat the steps for all the image blocks

2 (a): Read the value of the Threshold-cluster

2 (b): If data pixel-clusters ≥ Threshold-cluster

Make the image block as background block

End If

Step 3: End

More number of data pixel-clusters in a particular image block signifies that the block contains more noisy pixels. To remove these, a user defined parameter called Threshold-cluster is introduced. The performance of the post-processing method is heavily dependent on this parameter. The value of this parameter is tuned based on the nature of the document to yield better results.

4 Experimental Results of the Proposed Algorithm:

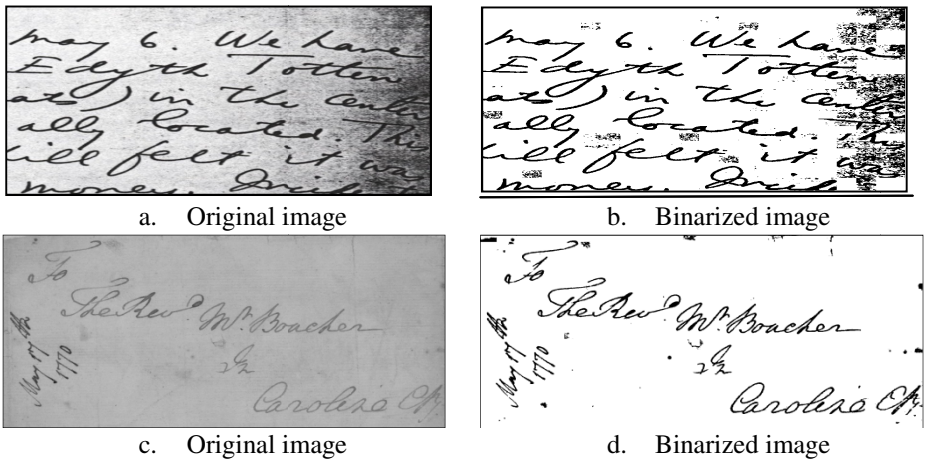
To evaluate the proposed binarization technique, document images are collected from different sources [12-14]. Both the good quality and degraded document images have been used to measure the strength of the algorithm. Different block sizes have been used for the *CDI* and the optimal result was found for the block size 30X30. Both the values of B_t and Threshold-cluster are determined experimentally. B_t is set to 30 and Threshold-cluster is set to 10 which produce optimum results for most of the cases.

The proposed technique produces satisfactory results on most of the document images. Some of the sample outputs of *SDI* and *CDI* are shown in Fig. 3 and Fig. 4 respectively.

a. Original image b. Binarized image

c. Original image d. Binarized image

Fig. 2. Binarized output images (b and d) of the original images (a and c) which are classified as *SDI*



a. Original image b. Binarized image

c. Original image d. Binarized image

Fig. 3. Binarized output images (b and d) of the original images (a and c) which are classified as *CDI*

5 Conclusion

Binarization is the process of converting a color or grayscale image into a bi-level image. The challenge lies in the task of determining the optimal threshold value(s) so that no background or data pixels are misclassified. In the present work, input document images are first categorized into 1) *SDI* and 2) *CDI*. Otsu's methodology has been directly applied to the SDI. To improve the resultant image quality, a post-processing is performed on the output image to retrieve the missing data pixels. Whereas for *CDI*, the image is divided into NxN blocks and local threshold values are determined using Otsu's algorithms. Post processing of document image is performed by detecting and deleting the noisy data blocks. The efficiency of the proposed algo-

rithm can be improved if the block size of the image and the value of the Threshold-cluster can be set dynamically. However, the present binarization algorithm is producing satisfactory results for a reasonable number of degraded documents.

References

1. Valizadeh, M., Armanfard, N., Komeili, M., Kabir, E.: A novel hybrid algorithm for binarization of badly illuminated document images. In: 14th International CSI Computer Conference (CSICC), pp. 121–126 (2009)
2. Kawano, H., Oohama, K., Maeda, H., Okada, Y., Ikoma: Degraded document image binarization combining local statistics. In: ICROS-SICE International Joint Conference, August 18-21 (2009)
3. Chang, Y.F., Pai, Y.T., Ruan, S.J.: An efficient thresholding algorithm for degraded document images based on intelligent block detection. In: IEEE Int. Conf. Syst. Man Cybern. SMC (2008)
4. Gatos, B., Pratikakis, I., Perantonis, S.J.: Efficient binarization of historical and degraded document images. In: The Eighth IAPR Workshop on Document Analysis Systems (2008)
5. Ostu, N.: A thresholding selection method from gray-level histogram. IEEE Trans. Systems Man Cybernet. SMC 8, 62–66 (1978)
6. Kittler, J., Illingworth, J.: Minimum error thresholding. Pattern Recognition 19(1), 41–47 (1986)
7. Bernsen, J.: Dynamic thresholding of gray-level images. In: Proc. Eighth International Conference on Pattern Recognition, Paris, pp. 1251–1125 (1986)
8. Niblack, W., Prentice, N.J., Cliffs, E.: An Introduction to Digital Image Processing (1986)
9. Sauvola, J., Pietikainen, M.: Adaptive Document Image Binarization. Pattern Recognition 33, 225–236 (2000)
10. Su, B., Lu, S., Tan, C.L.: Combination of Document Image Binarization Techniques. In: International Conference on Document Analysis and Recognition (2011)
11. Shaikh, S.H., Maiti, A.K., Chaki, N.: A new image binarization method using iterative partitioning. Springer (2012) (published online: January 6, 2012)
12. http://users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark/
13. http://users.iit.demokritos.gr/~bgat/H-DIBCO2010/resources.html
14. https://www.google.co.in/imghp?hl=en&tab=wi

Colored Rubber Stamp Removal from Document Images

Soumyadeep Dey*, Jayanta Mukherjee, Shamik Sural, and Partha Bhowmick

Indian Institute of Technology, Kharagpur
{soumyadeepdey@sit,jay@cse,shamik@sit,pb@cse}.iitkgp.ernet.in

Abstract. Rubber stamps on document pages often overlap and obscure the text very badly, thereby impairing its readability and deteriorating the performance of an optical character recognition system. Removal of rubber stamps from a document image is, therefore, essential for successfully converting a document image into an editable electronic form. We propose here an effective technique for rubber stamp removal from scanned document images. It is based on the novel idea of a single feature obtained by projecting the pixel colors of the image foreground along the eigenvector corresponding to the first principal component in HSV color space. Otsu's adaptive thresholding is used to segment out the stamp impressions from the text by exploiting the discriminative power of the aforesaid feature. Experimentation and subjective evaluation on a variety of scanned document images demonstrate the strength and effectiveness of the proposed technique.

Keywords: Rubber stamp removal, document cleaning, colored document processing.

1 Introduction

Rubber stamps, also called seals, are used to cast distinctive and lasting impressions on document pages. Their purpose is to certify a document for various reasons, such as authorization, authentication of source, etc. The seal essentially comprises a suitably molded or engraved 'pattern'. A usual practice before pressing the seal against a document page is to smear its 'pattern' with a specially made ink or dye, so that the required impression is properly transferred to the concerned page.

While scanning the stamp-containing pages, the stamp impressions also get scanned along with the actual data content, which poses severe problem in converting a document into an electronic text form by an optical character recognition (OCR) system. A stamp may be present in a page either in some blank space or in overlap with a text segment. The latter case is more problematic, since the performance of an OCR system falls drastically in presence of text-overlapping stamp regions. Our work is particularly focused on removing the text-overlapping

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 545–550, 2013.

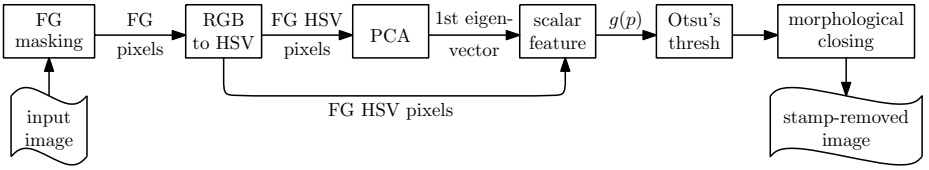


Fig. 1. Proposed stamp removal algorithm (FG = foreground)

stamp impressions from the concerned document image without affecting its text content. The OCR performance improves significantly once the stamp impressions are removed from the document image, as shown in this paper.

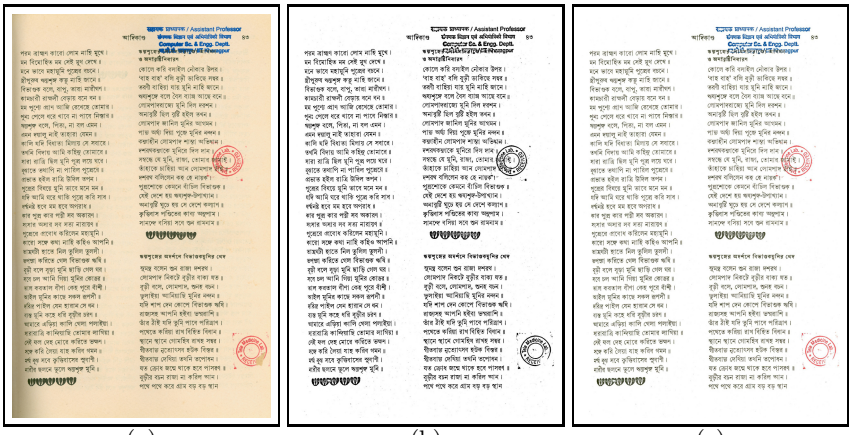
Till date, no work has been reported that addresses rubber stamp removal from document images, although there are some literature regarding rubber stamp detection in document images. In the stamp detection approach by Zhu and Doermann [7], stamps are limited to oval shape only. In another work, Zhu *et al.* [6] proposed an automatic logo detection algorithm, using a boosting strategy across multiple image scales. Forczmanski *et al.* [3] proposed a stamp detection algorithm based on color profile and shape analysis. All these methods are not appropriate for rubber stamp removal from document images, especially when the stamps do overlap with the text part. This has motivated us in designing an effective technique to remove rubber stamp impressions from document images. The work in this paper explicates this technique, which is based on principal component analysis in HSV space. One of its premises is that the text is written in a particular color of ink on a uniform background, and the stamps are of different color(s).

2 Methodology

The stages required for rubber stamp removal from document images are shown in Fig. 1. The input image is first classified into foreground and background. The foreground pixels are taken into HSV color space and PCA is performed on the converted foreground pixels. The foreground pixels in three-dimensional color space are then mapped onto an one-dimensional space, using the eigenvector corresponding to the first principal component. The obtained foreground pixels are classified into stamp and text pixels by usual thresholding technique after doing the histogram analysis of the resultant image.

2.1 Foreground Masking

An input color image I is first converted to a gray-scale image, I_g. The gray-scale image is binarized using an adaptive image binarization technique with window size 100×100 [2]. Binarization of the gray-scale image is obtained by Eq. 1, where $I_g(x, y)$ and $I_b(x, y)$ indicate the respective gray value and binary value at a pixel with coordinates (x, y) of the sub-image, and T represents the dynamically obtained threshold value for the sub-image.



(a) (b) (c)

Fig. 2. (a) An input image; (b) binary image; (c) foreground masked image

$$I_b(x,y) = \begin{cases} 255 & \text{if } I_g(x,y) > T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Let F and B denote the respective sets of foreground pixels and of background pixels. The set F is used for masking the source color image on the binary image. This masking operation is used to get back the color information of the foreground pixels from the source image. The image, thus formed, is known as *foreground masked image*, has a uniform background and is used for further processing. An example of sample input image, its corresponding binary image, and foreground masked image are shown in Fig. 2.

2.2 Stamp Removal and Output Image Generation

The foreground masked image is an image in RGB color space. The foreground pixels of the foreground masked image are converted from RGB to HSV color space [4]. In HSV color space, a feature is selected for segmentation of stamp from the text part.

Principal component analysis (PCA) is a mathematical procedure to convert a set of correlated variables into a set of linearly uncorrelated variables called principal components, using an orthogonal transformation. Principal component analysis is carried out by computing the eigenvectors of the 3×3 covariance matrix in HSV space. We consider a feature based on the first principal component of the foreground data. Hence, the corresponding *unit eigenvector* is computed. Let the obtained unit eigenvector be $\overrightarrow{u} = a\hat{i} + b\hat{j} + c\hat{k}$. Let us denote the color vector at pixel p as $\overrightarrow{f(p)} = h(p)\hat{i} + s(p)\hat{j} + v(p)\hat{k}$, where $h(p)$, $s(p)$, and $v(p)$ denote the hue, saturation, and intensity values of p in HSV color space. Then the foreground three-dimensional data is converted into one-dimensional (scalar) data along the first principal component by projecting $\overrightarrow{f(p)}$ along $\overrightarrow{u}$ according to the following equation.

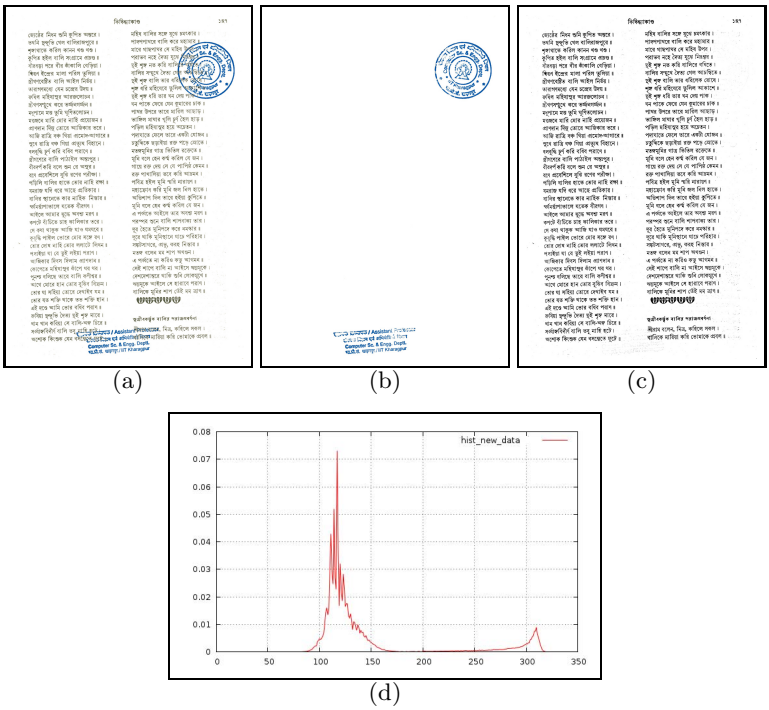


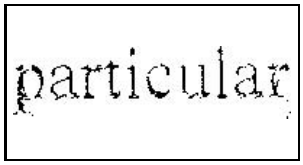
Fig. 3. (a) A foreground masked image; (b) stamp part of the image; (c) text part of the image after stamp removal; (d) histogram of foreground data ($g(p)$)

$$g(p) = \overrightarrow{f(p)} \cdot \overrightarrow{u} \tag{2}$$

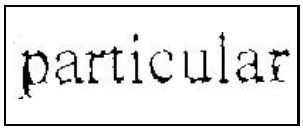
For segmentation, the histogram of $g(p)$ is computed and analyzed as a bimodal histogram. Its highest peak corresponds to the text part, and the other peak to the stamp region(s). Otsu's thresholding algorithm [5] is used to classify the foreground pixels into two classes, one representing the text region and the other representing the stamp region.

Figure 3 shows a result produced by our algorithm. The foreground masked image is shown in Fig. 3(a). The histogram obtained using Eq. 2 on the foreground pixels is shown in Fig. 3(d). The segmented stamp regions and the text regions after segmenting the foreground pixels, using Otsu's thresholding method on the obtained histogram data, are shown in Fig. 3(b) and Fig. 3(c) respectively.

To join the characters, which got broken due to stamp removal from overlapped regions, morphological closing is performed by applying a dilation operation followed by an erosion [4]. The objective of this operation is to fill the holes in the text components without distorting their boundaries as much as possible. One example of a word with broken characters and its closed image obtained by using 3×3 morphological kernel, along with their corresponding OCR output, are shown in Fig. 4.



(a) OCR output = "jparticulasz"



(b) OCR output = "particular"

Fig. 4. (a) Binary image and its OCR output; (b) image after morphological closing and its OCR output

3 Results

For evaluation of the proposed rubber stamp removal algorithm, experimentation is conducted on different scanned pages. These pages are in different languages and contain rubber stamps of different shapes and colors. All the pages are scanned in 300 dpi resolution and given as input to the rubber stamp removal system. The input images and stamp-removed images are used for evaluation. The documents are read one at a time. The number of words readable before stamp removal and that after stamp removal are counted for the evaluation purpose. For documents written in English language, the OCR performance is reported before and after rubber stamp removal using a standard OCR system [1]. The OCR performance is not shown for Bengali documents, as its performance on our Bengali data set is very poor.

In Table 1, reading performances for stamp-overlapping text regions are shown, before and after stamp removal from the document images. Here, n_w refers to the number of words overlapping with the stamp regions, and Acc_{ws} and Acc_{sr} denote the respective accuracies for stamp-overlapping text regions before and after stamp removal. It may be noticed from this table that the reading accuracies improve significantly for both English and Bengali document images, once the stamp impressions are removed. The OCR performance also has an encouraging improvement for the English documents, as reflected in the corresponding word-level and character-level accuracies shown in Table 2. The accuracies of the OCR on English words and characters are estimated by manually counting the number of words and their associated characters which are overlapped with stamps. Some sample stamp-overlapping text regions and their outputs after stamp removal are shown in Fig. 5.

Table 1. Reading performance for English and Bengali documents

document	n_w	$Acc_{ws}(\%)$	$Acc_{sr}(\%)$
Bengali	264	28.41	98.11
English	221	39.82	99.09

Table 2. OCR performance on English documents

data type	number	$Acc_{ws}(\%)$	$Acc_{sr}(\%)$
words	221	1.81	84.62
characters	958	18.68	91.96

Fig. 5. Sample stamp removal results

4 Conclusion

We have proposed a stamp removal technique from document images, which significantly improves the document reading accuracy, as evidenced by relevant experimentation. The OCR performance for English documents in particular is found to be improved to a significant extent. The proposed technique can, therefore, be included in the document cleaning stage to improve the performance of an OCR system.

Acknowledgment. This work is funded by TCS research scholar program and Ministry of Communications & Information Technology, Government of India, Ref.: MCIT 11(19)/2010 HCC (TDIL) dt. 28-12-2010. We are thankful to our colleagues and lab members for preparing the data sets.

References

1. Free online OCR, `http://www.newocr.com/`
2. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
3. Forczmański, P., Frejlichowski, D.: Robust stamps detection and classification by means of general shape analysis. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2010, Part I. LNCS, vol. 6374, pp. 360–367. Springer, Heidelberg (2010)
4. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. PHI (2009)
5. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. SMC 9(1), 62–66 (1979)
6. Zhu, G., Doermann, D.: Automatic document logo detection. In: Proc. ICDAR 2007, pp. 864–868 (2007)
7. Zhu, G., Jaeger, S., Doermann, D.: A robust stamp detection framework on degraded documents. In: SPIE Conf. Doc. Recog. & Retrieval, pp. 1–9 (2006)

On the Role of Compensatory Operators in Fuzzy Result Merging for Metasearch

Arijit De

Tata Consultancy Services, Mumbai, India
arijit.axd9142@gmail.com

Abstract. A key metasearch engine task is result merging of search results from multiple search engines in response to a user query. The problem of result merging has been widely studied as a multi-criteria decision making model (MCDM). While many MCDM techniques have been employed to create experimental models for result merging, the most notable have used fuzzy aggregation operators such as the OWA operators and its extensions and variations. In this work we study the role of applying fuzzy algebraic t-norms, s-norms and compensatory operators in fuzzy result merging for metasearch. Our results will demonstrate the superiority of compensatory operators over t-norm aggregation functions in the context of result merging for metasearch.

Keywords: Information Retrieval, Metasearch Engines, Fuzzy Sets, Fuzzy Aggregation Operators.

1 Introduction

A metasearch engine is a search engine that can be used to search multiple search engines systems concurrently. Metasearch engines are particularly useful in searching through topic specific search systems like PUBMED, MEDILE etc. A typically search engine output comprises of a list of results (documents/URLs/database records), ranked in the order of relevance. However, different search engines evaluate and consequently rank results differently. The problem of result merging is to aggregate the rankings of each result to come up with a composite rank, such that the final ranking preserves the order of relevance. In this paper we propose two models Compensatory Ordered Weights Average (COWA) and Importance Guided Compensatory Ordered Weights Average (IGOWA) models for result aggregation using the OWA operator [11] and compensatory aggregation function [15]. We compare our models with the existing OWA [5], IGOWA, t-norm OWA and t-norm IGOWA [3, 4]. This paper is organized as follows. In Section 2, we discuss previous models for result merging, including a discussion on the OWA, IGOWA, t-norm OWA and t-norm IGOWA models. In Section 3 we describe our proposed models, COWA and IGCOWA. In Section 4 we describe experiments comparing COWA and IGCOWA with OWA, IGOWA and t-norm IGOWA and discuss our results. In Section 5 summarize our findings in a conclusion.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 551–556, 2013.

2 Previous Work

In early work on aggregation, includes the work of Fox and Shaw [6, 7] and Aslam and Montague [1]. The latter proposed two models Borda-Fuse and Weighted Borda-Fuse based on the political election strategy, Borda Count [2].

Diaz [5] developed the first fuzzy result aggregation model OWA, based on Yager Ordered Weighted Average (OWA) [11, 12] operator. The OWA model uses a *positional value* (PV) to quantify the rank of a result in a result list. The positional value (PV) of a result ranked r in a result list is (n – r + 1) where n is the total number of results in the list. The OWA model uses the OWA operator to aggregate the PVs of each result. Let us say we have n criteria and an alternative x. Let a_i represent the degree to which x satisfies the i^{th} criteria. Thus we have a set $\{a_1 \ldots a_n\}$. Let b_j is the j^{th} largest value within the set $\{a_1, a_2, \ldots, a_n\}$. Then F (eq. 1) defines the OWA operator.

$$F(a_1, a_2, a_3, \ldots, a_n) = \sum_{j=1}^{n} w_j b_j \tag{1}$$

In the OWA model for metasearch, the PV (or inverse of rank) is considered extent to which the result (alternative) satisfies a search engine (criteria). The ordered weights are computed using a linguistic RIM quantifier $Q(r) = r^{\alpha}$ as described in equation (2). The orness associated with the quantifier, orness $(Q) = (1+\alpha)^{-1}$. Experiments in [10, 11] demonstrate the OWA model outperforms the Borda Fuse and Weighted Borda Fuse models.

$$w_i = Q\left(\frac{i}{n}\right) - Q\left(\frac{i-1}{n}\right) \tag{2}$$

The OWA model however, does not consider search engine importance weights in result aggregation. To overcome this, De [3, 4] proposed the IGOWA (importance guided) model for metasearch. The IGOWA model uses Yager's [12] Importance Guided Aggregation technique to generate the ordered weights for aggregation using the OWA operator. Once again PVs of results from result lists are aggregated using the OWA operator. However, weights are generated as per equation (3) below. Let the importance weight for the i^{th} criteria be V_i. Let $V_i \in [0, 1]$. For an alternative x there will be a pair (V_i, a_i) for each criteria i. The criteria scores can be sorted in descending order with b_k being the k^{th} largest a_i. Let u_k be the importance weight attached to b_k. We can now associate, with alternative x, a collection of n (u_k, b_k) pairs, where the b_k's are degrees to which x satisfies the n criteria in descending order. The ordered weights can now be obtained using equation (3) and where $T = \sum_{j=1}^{n} u_j$ .Yager [12] proposes a set of extensions of the OWA operator, which he calls the triangular norm (t-norm) OWA operators. This is essentially a quantifier guided aggregation decision function that strikes a balances the RIM quantifier Q, defined previously, that stipulates the degree of satisfaction that is attained when satisfying i^{th} criteria with the need to find i criteria that are satisfied. Combining these two factors, Yager arrives at the aggregation function shown in eq. (4). Here b_j is j^{th} greatest a_i and T is a t-norm function.

$$w_k(x) = Q\left(\frac{\sum_{j=1}^{k} u_j}{T}\right) - Q\left(\frac{\sum_{j=1}^{k-1} u_j}{T}\right) \tag{3}$$

$$F(a_1, a_2, a_3, ..., a_n) = \sum_{j=1}^{n} w_j T(b_1, .., b_j) \tag{4}$$

De [3, 4] used an algebraic t-norm eq. (5) to propose two models for result merging. The first was the t-norm OWA result merging model for metasearch and did not consider search engine importance weights. The second was the importance guided t-norm IGOWA model that use search engine weights to generate ordered weights using equation (3). In each case multiple PVs for each result (obtained from different search engine result lists) are obtained. Experiments of De [3, 4] demonstrated that when using an algebraic t-norm the t-norm OWA model outperformed the OWA model and the t-norm IGOWA model outperformed the IGOWA model for metasearch. The IGOWA and t-norm OWA model both outperformed the OWA model in creating improved relevance order ranking.

$$T(b_1, ...b_n) = \prod_{i=1}^{n} b_i \tag{5}$$

3 Proposed Models

Several researchers working in the area of fuzzy decision making noticed that t-norms and their dual s-norms lack the compensation behavior crucial to any aggregation process. Zimmermann and Zysno [15] were the first to notice that in a decision function making context humans neither follow the behavior of a t-norm or an s-norm strictly when aggregating. Zimmermann and Zysno [15] proposed an aggregation function on the unit interval based on t-norms and s-norms as described in equation (6). Here γ is the extent of compensation provided. Yager [14] proposes a function to calculate the value of γ as described in equation (7).

The motivation of our work was to study how using compensatory operators, affects the result of merging result merging for metasearch. We build two models. Our first model was the Compensatory Ordered Weighed Aggregation (COWA) model. This model employs the Zimmermann [15] compensatory aggregation function in result aggregation using the OWA operator as defined in eq. (8). This model does not take into consideration search engine weights in result merging. In our model we use the algebraic t-norm for aggregation as described in eq. (5). We compare the performance of this model to the OWA model [5] and t-norm OWA model [3, 4].

Our second model is the Importance Guided Compensatory Ordered Weighted Aggregation (IGCOWA) model for result merging. The model is similar as it user the

Zimmerman [15] aggregation function in conjunction with the OWA operator as described in eq. (8). However the ordered weights for aggregation are generated using equation (3) as in IGOWA and t-norm IGOWA [3, 4]. Both our models use a Regular Increasing Monotone (RIM) quantifier as described earlier of the form $Q(r) = r^{\alpha}$ as a function to generate ordered weights.

$$Z_{\gamma}(b_1, b_2, b_3, ..., b_n) = \left(\prod_{i=1}^{n} b_i\right)^{1-\gamma} \bullet \left(1 - \prod_{i=1}^{n} (1 - b_i)\right)^{\gamma} \tag{6}$$

$$\gamma = \frac{T(b_1, ..., b_n)}{T(b_1, ..., b_n) + T(1 - b_1, ..., 1 - b_n)} \tag{7}$$

$$F(a_1, a_2, a_3, ..., a_n) = \sum_{j=1}^{n} w_j Z_{\gamma}(b_1, .., b_j) \tag{8}$$

4 Experiments and Results

For our experiments, we use Hersh's [8] OHSUMED collection within the LETOR 2 (**L**earning **TO R**ank) [9] dataset from Microsoft Research Asia. The LETOR2 dataset comprises of a collection of OHSUMED documents (results), a query set of size of 106, a set of 25 algorithms (search engines) that are used to judge the relevance of each document to each query. For our experiments, our performance metric is Recall-Based (RB) Precision as defined by Bollmann and Raghavan [16]. We compare our proposed models COWA and IGCOWA against existing models OWA, t-norm OWA, IGOWA and t-norm IGOWA. Odd numbered queries are used for learning search engine importance weights based on performance of search engines over the query set. For our experiments for each odd numbered query, we randomly pick N search engines from the 25 available. The value of N is varied from 2 to 12. Overall 1000 sets of experiments are done for each value of N. Table 1 shows the results. We compute the average precision at recall levels of 0.25, 0.5, 0.75 and 1. The results are shown in Table 1 and Table 2.

Table 1. Results comparing COWA, IGCOWA vs. OWA, IGOWA, t-norm OWA and t-norm IGOWA when number of lists varies from 2 to 12.

N (Number of lists merged)	Average Precision of the Merged List					
	OWA	IGOWA	t-norm OWA	t-norm IGOWA	COWA	IGCOWA
2	0.4051	0.4231	0.4233	0.4472	0.4538	0.4638
4	0.4237	0.4445	0.4453	0.4491	0.4573	0.4783
6	0.4297	0.4593	0.4597	0.4638	0.4791	0.4891
8	0.4332	0.4682	0.4712	0.4783	0.5011	0.5013
10	0.4681	0.4783	0.4813	0.4891	0.5113	0.5291
12	0.4732	0.4813	0.4913	0.5013	0.5231	0.5345

Table 2. Results comparing COWA, IGCOWA vs. OWA, IGOWA, t-norm OWA and t-norm IGOWA when orness of aggregation varies from 0.8 to 0.2.

Orness (O)	Average Precision of the Merged List					
	OWA	IGOWA	t-norm OWA	t-norm IGOWA	COWA	IGCOWA
$O \geq 0.8$	0.4371	0.4413	0.4417	0.4472	0.5146	0.5292
$0.8 \geq O \geq 0.6$	0.4251	0.4345	0.4453	0.4491	0.4629	0.5177
$0.6 \geq O \geq 0.4$	0.4108	0.4139	0.4397	0.4428	0.4439	0.5237
$0.4 \geq O \geq 0.2$	0.4332	0.4428	0.4712	0.4783	0.5211	0.5378

5 Conclusions

From Table 1 it is clear that the COWA model outperforms the models in its class (without search engine importance weights), namely the OWA and the t-norm OWA (with algebraic t-norms) models. The overall improvements over t-norm OWA model are 7.20, 2.69, 4.22, 6.34, 6.23 and 6.47 % when merging 2, 4 6, 8, 10 and 12 search engines respectively. Similarly, in its class of models (requiring search engine importance weights) IGCOWA model improves upon the IGOWA model by 3.7, 6.5, 5.45, 4.80, 8.17 and 6.62 % when merging 2, 4, 6, 8, 10 and 12 search engines. . In Table 1 it is also observed that as the number of search engines increases the overall performance in terms of average precision increases. This is primarily because when more search engine results are merged more relevant results are present in the merged list.

Orness is a key measure in fuzzy aggregation. For the OWA, t-norm OWA and COWA models orness is computed as $1/(1+\alpha)$. Here α is the parameter of the RIM quantifier $Q(r) = r^{\alpha}$ used to compute ordered weights. We also measured the performance of aggregation for different levels of orness. In previous findings [3,4, 5] it was observed that as orness of aggregation decreases, the performance in terms of average precision decreases, till averaging conditions are achieved. When orness =0.5 then the average precision falls to its lowest value. Following this as orness decreases and andness increases the average precision increases.

This is consistent with findings in [3, 4 and 5] illustrated in results from Table 2. Also for each level of orness the performance of COWA is better than that of algebraic t-norm OWA and OWA. For the models that require importance weights the IGCOWA model also outperforms the t-norm IGOWA and IGOWA models.

The overall improvements of COWA over OWA and t-norm OWA are 9.98% and 8.01%. The overall improvements of IGCOWA over IGOWA and t-norm IGOWA are 5.52% and 5.87%. Clearly compensatory operators clearly improve the performance of aggregation in terms of average precision.

References

1. Aslam, J., Montague, M.: Models for metasearch. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2001), New Orleans, LA, USA, September 1-6, pp. 276–284. ACM Press, New York (2001)
2. Borda, J.C.: Memoire sur les elections au scrutiny. Histoire de l'Academie Royale des Sciences, Paris (1781)
3. De, A., Diaz, E.D., Raghavan, V.V.: On Fuzzy Result Merging for Metasearch. In: IEEE International Conference on Fuzzy Systems, London, United Kingdom. IEEE Press (July 2007)
4. De, A.: Fuzzy result merging models for metasearch, A Dissertation. University of Louisiana Press (2006)
5. Diaz, E.D., De, A., Raghavan, V.V.: A comprehensive OWA-based framework for result merging in metasearch. In: Ślęzak, D., Yao, J., Peters, J.F., Ziarko, W.P., Hu, X. (eds.) RSFDGrC 2005. LNCS (LNAI), vol. 3642, pp. 193–201. Springer, Heidelberg (2005)
6. Fox, E.A., Shaw, J.A.: Combination of multiple searches: Part 2. In: Proceedings of the 2nd Text Retrieval Conference (TREC-2), Gaithersburg, MD, August 31-September 2, pp. 243–252. US Government Printing Office, Washington DC (1993)
7. Fox, E.A., Koushik, M.P., Shaw, J.A., Modlin, R., Rao, D.: Combination of multiple searches: Part 1. In: Proceedings of the 1st Text Retrieval Conference (TREC-1), Gaithersburg, MD, USA, November 4-6, pp. 319–328. US Government Printing Office, Washington DC (1992)
8. Hersh, W., Buckley, C., Leone, T.J., Hickam, D.: OHSUMED: An interactive retrieval evaluation and new large test collection for research. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 192–201. ACM/Springer, New York, NY (1994)
9. Liu, T., Xu, J., Qin, T., Xiong, W., Li, H.: LETOR 2007: Benchmark dataset for re-search on learning to rank for information retrieval. In: LR4IR 2007, in Conjunction with SIGIR 2007 (2007)
10. Turksen, I.B.: Interval-valued fuzzy sets and 'compensatory AND'. Fuzzy Sets and Systems 51, 295–307 (1992)
11. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. Fuzzy Sets and Systems 2, 243–260 (1983)
12. Yager, R.R.: Quantifier guided Aggregating using OWA operators. International Journal of Intelligent Systems 1, 49–73 (1996)
13. Yager, R.R., Kreinovich, V.: On how to merge sorted lists coming from different web search tools. Soft Computing Research Journal 3(1), 83–88 (1999)
14. Yager, R.R., Rybalov, A.: Full reinforcement operators in aggregation techniques. IEEE Transactions on Systems, Man and Cybernetics 28, 757–769 (1998)
15. Zimmermann, H.-J., Zysno, P.: Latent connectives in human decision making. Fuzzy Sets and Systems 4, 37–51 (1980)
16. Bollmann, P., Raghavan, V.V., Jung, G.S., Shu, L.C.: On probabilistic notions of precision as a function of recall. In: Information Processing and Management, pp. 291–315. Pergamon Press, Tarrytown (1992)

The Fuzzy Parametrized Model for Classifying Blocks in the Non-binary Motion Mask

Dmitry A. Matsypaev[1] and Andrey G. Bronevich[2]

[1] Southern Federal University,Taganrog Institute of Technology, Taganrog, Russia
[2] National Research University "Higher School of Economics", Moscow, Russia
{dmitry.matsypaev,andreybronevich}@gmail.com

Abstract. The motion detection in video is considered. We break non-binary motion mask on blocks and calculate a certain statistics for each block. Then we use prior information about statistics distribution to classify blocks on background and foreground. The estimation framework for classification confidence is presented.

Keywords: non-binary motion mask, block representation, fuzzy model, imprecise probabilities.

1 Introduction

This paper tackles the problem of automatic detection of moving objects in the video sequence taken from the stationary monocular camera with the fixed focal length. Existing methods to solve it are described in detail in [1],[2],[5],[7],[8], [10],[11] etc.

We propose to build non-binary motion mask and to break it on rectangular blocks. For each block we estimate the special statistics whose distribution allows us to decide, whether a certain block belongs to foreground or to background. We introduce the fuzzy mathematical model of classification that is based on the prior information of how moving objects appear in the video sequence.

2 Building the Motion Mask

Let us assume that we have the sequence of video frames $\{F^{(k)}\}$, where each frame is a gray scale image with the $W_F \times H_F$ resolution. In [1],[2],[5],[7],[8],[10],[11] the motion mask $M^{(k)}$ for the frame $F^{(k)}$ is a binary image derived from the frame set $\{F^{(k-m+1)}, \ldots, F^{(k)}\}$, $m > 1$. To build a binary mask, the threshold transformation can be used. Applied to the signal, we can loose some useful information about motion objects, therefore, in the paper we suggest to use the non-binary motion mask. The simplest way to obtain it is to compute the absolute difference between two neighboring frames without threshold transformation:

$$M_{ij}^{(k)} = |F_{ij}^{(k)} - F_{ij}^{(k-1)}| \ , \tag{1}$$

where $1 \le i \le W_F$, $1 \le j \le H_F$, $0 \le M_{ij}^{(k)} \le I_{max}$, and I_{max} is the maximal value of the pixels intensity.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 557–562, 2013.

3 Representation of Blocks in the Motion Mask

The representation of images with the help of blocks is widely used in video coding [9] and image processing[4], [8]. In this case we break each motion mask on rectangular blocks with constant size $W_B \times H_B$. Let U be the set of such blocks. For each block $B \in U$ the following statistics is computed:

$$I(B) = \frac{1}{W_B H_B} \sum_{i=1}^{W_B} \sum_{j=1}^{H_B} B_{ij}, \tag{2}$$

where B_{ij} is the pixel intensity value for block B at column i and row j. Such a motion mask representation fits the localization and compactness properties of moving areas, reduces computational complexity and decreases the effect of noise. However, the geometrical precision of the result is bounded from below by the size of the block.

4 Foreground Blocks Detection

4.1 Basic Hypotheses

We will describe next the properties of motion blocks with the help of two fuzzy sets: by a fuzzy set of foreground blocks U_{fg} with the membership function $\mu_{fg}(B)$ and by a fuzzy set of background blocks U_{bg} with the membership function $\mu_{bg}(B)$, where $B \in U$. Let us assume that these sets obey the following hypothesis.

Hypothesis 1. For all $B_1, B_2 \in U$ the statement $I(B_1) \geq I(B_2)$ is true if and only if $\mu_{bg}(B_1) \leq \mu_{bg}(B_2)$ and $\mu_{fg}(B_1) \geq \mu_{fg}(B_2)$.

The Hypothesis 1 states that the statistics (2) can be used to order the elements of the set U by the membership functions values. Assuming that the Hypothesis 1 is true, let us postulate the additional two hypotheses.

Hypothesis 2. Let $C_{bg} = \{B : \mu_{bg}(B) = 1\}$. Then $|C_{bg}|/|U| = \alpha$, where $\alpha \in (0, 1)$ is the parameter of the hypothesis.

Hypothesis 3. Let $C_{fg} = \{B : \mu_{fg}(B) = 1\}$. Then $(I_{max} - \min_{B \in C_{fg}} I(B))/I_{max} = \beta$, where $\beta \in (0, 1)$ is the parameter of the hypothesis.

Hypotheses 2 and 3 are derived from the analysis of a typical plot of the cumulative distribution function computed for the statistics (2)

$$F_I(x) = \frac{1}{|U|} \sum_{B \in U | I(B) \leq x} 1 \ .$$

As you can see on Fig. 1, for the most of blocks the statistics I is close to zero while for the minority of blocks I is dispersed over the segment $[0, I_{max}]$. It means that both α and β should be close to 1. We set these parameters to 0.9 in our experiments.

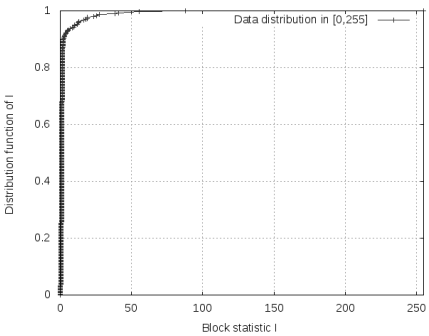


Fig. 1. The typical plot for F_I

4.2 Membership Function Formalization

Let us introduce the value $I_\alpha = \min_{x\in[0,I_{max}]}\{x|F_I(x) \geq \alpha\}$. Then the formalization of membership function $\mu_{bg}(B)$ based on Hypotheses 1 and 2 leads to the following. Let $\mu_{bg}(B) = \eta_{bg}(I(B))$, where $\eta_{bg}(x)$ satisfies the following conditions: 1) $\eta_{bg}(x)$ is defined on $[0, I_{max}]$, 2) $\forall x \in [0, I_\alpha]$ $\eta_{bg}(x) = 1$, 3) $\eta_{bg}(I_{max}) = 0$ and 4) $\eta_{bg}(x)$ is monotone and non-increasing on $[0, I_{max}]$.

Analogously, let us introduce the value $I_\beta = (1-\beta)I_{max}$. Then the formalization of membership function $\mu_{fg}(B)$ based on Hypotheses 1 and 3 leads to the following. Let $\mu_{fg}(B) = \eta_{fg}(I(B))$, where $\eta_{fg}(x)$ satisfies the following conditions: 1) $\eta_{fg}(x)$ is defined on $[0, I_{max}]$, 2) $\forall x \in [I_\beta, I_{max}]$ $\eta_{fg}(x) = 1$, 3) $\eta_{fg}(0) = 0$ and 4) $\eta_{fg}(x)$ is monotone and non-decreasing on $[0, I_{max}]$.

4.3 Tversky's Functional

The parametrized Tversky's functional is used in fuzzy set theory to estimate the degree of similarity of two sets [3]:

$$T_{a,b}(X,Y) = \frac{f(X\cap Y)}{f(X\cap Y) + af(X-Y) + bf(Y-X)} \quad ,$$

where X,Y is a pair of fuzzy sets, $\cup$, $\cap$, $-$ are consistent operations to unite, intersect and subtract two fuzzy sets, f is a functional that satisfies $f(X\cup Y) = f(X) + f(Y)$ for all disjoint sets X, Y, and $a, b \in \mathbb{R}$ are parameters.

It is possible to prove that functions η_{bg} and η_{fg} satisfy the conditions proposed above if they are defined as

$$\eta_{bg}(x) = \Big(\phi(x, I_\alpha) - \phi(I_{max}, I_\alpha)\Big) \Big/ \Big(1 - \phi(I_{max}, I_\alpha)\Big),$$

where

$$\phi(x,y) = T_{1/2,0}\left([-\gamma/2, x], [-\gamma/2, y]\right) = \frac{2\min\{x,y\} + \gamma}{2\min\{x,y\} + \gamma + \max\{x-y,\ \ 0\}},$$

and

$$\eta_{fg}(x) = \Big(\psi(x, I_\beta) - \psi(0, I_\beta)\Big) \Big/ \Big(1 - \psi(0, I_\beta)\Big),$$

where

$$\psi(x, y) = T_{0,1/2}\left([-\gamma/2, x], [-\gamma/2, y]\right) = \frac{2\min\{x, y\} + \gamma}{2\min\{x, y\} + \gamma + \max\{y - x,\ \ 0\}}.$$

The calibration parameter $\gamma > 0$ is chosen such that

1. $\phi(0, 0) = 1$;
2. for fixed $x_0 \in (0, I_{max}]$ and $\phi_0 \in (0, 1)$ the statement $\phi(x_0, 0) = \phi_0$ is true.

We used $x_0 = 1$ and $\phi_0 = 2/3$ in our experiments. If $I_{max} = 255$, then $\gamma = 2$.

4.4 Block Classification

To determine whether a block is foreground or background, we need to check inequality

$$\mu_{fg}(B) \geq \mu_{bg}(B) \quad . \tag{3}$$

We define below the probability $p_{fg}(B)$ of block $B \in U$ to be in the foreground using the model of imprecise probabilities [6]. In this case the lower probability is defined as $\underline{p}_{fg}(B) = \min\{\mu_{fg}(B), 1 - \mu_{bg}(B)\}$ and the upper probability is defined as $\overline{p}_{fg}(B) = \max\{\mu_{fg}(B), 1 - \mu_{bg}(B)\}$. Such estimates are not contradictory because $[\underline{p}_{fg}(B), \overline{p}_{fg}(B)]$ is a convex set[6].

We take a mean point of the set above as the expected foreground probability value $p_{fg}(B)$. Obviously,

$$p_{fg}(B) = 0.5\left(\mu_{fg}\left(B\right) + 1 - \mu_{bg}\left(B\right)\right). \tag{4}$$

The equation (4) allows us to define a more general criterion comparing to (3):

$$p_{fg}(B) \geq p_{th}, \tag{5}$$

where $p_{th} \in [0, 1]$ is a threshold parameter. Let us notice that inequality (3) is equivalent to (5) if $p_{th} = 1/2$.

4.5 Confidence Estimation

The use of imprecise probabilities allows us to estimate the confidence of classification in bits. Let a block B be classified then we can estimate the distrust level using the formula below:

$$D(B) = -p_{med}\log_2 p_{med} - (1 - p_{med})\log_2(1 - p_{med}) \quad ,$$

where $p_{med} = median\{\overline{p}_{fg}(B), \underline{p}_{fg}(B), 1/2\}$. The distrust level depends on the block statistics I as shown on Fig.2. Its distribution is unimodal with a high peak where statistics I has intermediate values.

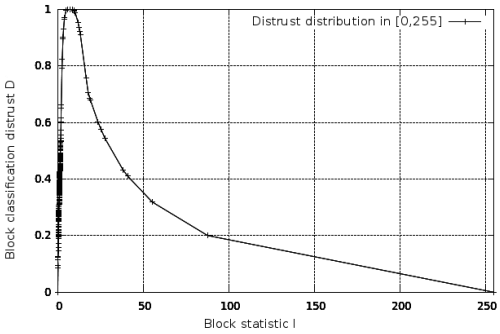


Fig. 2. The dependence between block statistics I and estimated distrust level D

5 Experiments

To proceed the quality evaluation, we have manually labeled 3 video sequences. During the experiments we set the size of blocks to 16×16 and the parameter p_{th} to 0.5 For the comparison we have chosen the Mixture of Gaussians (MoG) algorithm [11] implemented in the OpenCV programming library[2]. It is an ancestor of a lot of more specialized and slow state-of-the-art techniques which are adapted to specific cases [1].

On the Fig. 3 one can see the ROC-curve which compares the presented algorithm and the MoG. It should be noted that our algorithm bounds the precision from below because it operates with each the block containing 256 pixels as the whole. That's why MoG superiors in the $[0.13; 0.97]$ segment of the True Positive Rate. As for the performance evaluation, our algorithm is faster than MoG $2 - 5$ times depending on a certain video.

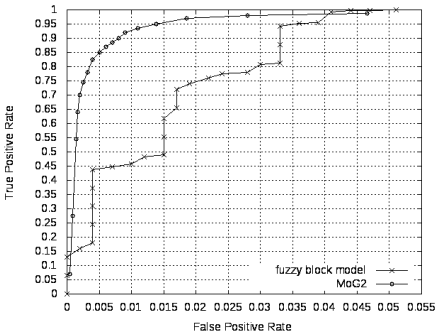


Fig. 3. The ROC-curve to compare presented algorithm (aka fuzzy block model) with Mixture of Gaussians (aka MoG2)

6 Conclusion

The approach of detecting moving objects in the video sequence is presented. The main idea is to construct the non-binary motion mask and to use its block representation for further analysis. The latter is based on the mathematical model that formalizes prior information and hypotheses obtained during the proceeded experiments. The mathematical framework to estimate the confidence of motion mask segmentation is considered. Further research will be dedicated to choosing the optimal parameters of the model presented above and to constructing moving areas from the foreground blocks.

References

1. Bouwmans, T.: Recent advanced statistical background modelling for foreground detection: a systematic survey. Recent Patents on Computer Science 4(3), 147–176 (2011)
2. Bradski, G., Kaehler, A.: Learning OpenCV. Computer vision with the OpenCV library. O'Reilly Media, Inc., Sebastopol (2008)
3. Cross, V.V., Sudkamp, T.A.: Similarity and compatibility in fuzzy set theory: assessment and applications. Physica-Verlag GmbH & Co., Heidelberg (2002)
4. Guaourova, A., Kamath, C., Cheung, S.C.: Block matching for object tracking. Technical report, Lawrence Livermore National Laboratory (2003)
5. Hofmann, M., Tiefenbacher, P., Rigoll, G.: Background segmentation with feedback: the pixel-based adaptive segmenter. In: Proc. of IEEE Workshop on Change Detection, pp. 38–43 (2012)
6. Klir, G.J.: Uncertainty and information. Foundations of generalized information theory. Jonh Wiley & Sons, Inc., New York (2006)
7. Narayana, M., Hanson, A.R., Learned-Miller, E.G.: Background modelling using adaptive pixelwise kernel variances in a hybrid feature space. In: CVPR, pp. 2104–2111. IEEE (2012)
8. Park, J., Tabb, A., Kak, A.C.: Hierarchical data structure for real time background subtraction. In: IEEE International Conference on Image Processing, pp. 1849–1852. IEEE, Atlanta (2006)
9. Richardson, I.E.G.: H.264 and MPEG-4 video compression: video coding for next generation multimedia. Jonh Wiley & Sons Ltd., Chischester (2003)
10. Van Droogenbroeck, M., Paquot, O.: Background subtraction: experiments and improvements for ViBe. In: Proc. of 2012 IEEE Computer Society Conference, pp. 32–37. Providence (2012)
11. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: International Conference Pattern Recognition, vol. 2, pp. 28–31. ICPR, Cambridge (2004)

Homomorphisms on the Monoid of Fuzzy Implications $(\mathbb{I}, \circledast)$ - A Complete Characterization

Nageswara Rao Vemuri and Balasubramaniam Jayaram

Department of Mathematics,
Indian Institute of Technology Hyderabad,
Yeddumailaram - 502 205, India
{ma10p001,jbala}@iith.ac.in

Abstract. In [4], we had proposed a novel generating methods of fuzzy implications and investigated algebraic structures on the set of all fuzzy implications, which is denoted by $\mathbb{I}$. Again in [5], we had defined a particular function g_K on the monoid $(\mathbb{I}, \circledast)$ (See Def. 16) and characterised the function K for which g_K is a semigroup homomorphism (s.g.h) in two special cases, i.e., K is with trivial range and $K(1, y) = y$ for all $y \in [0, 1]$(neutrality property). In this work we characterise the nontrivial range non neutral implications K such that g_K is an s.g.h. and also present their representations.

Keywords: Fuzzy implication, neutrality property, homomorphism.

1 Introduction

Fuzzy implications are a generalisation of classical implication to the fuzzy logic. It has many applications in the areas like approxiamte reasoning, control theory, decision making, fuzzy logic and so on. Its definition is as follows:

Definition 1. *[1] A binary function I on $[0, 1]$ is called a fuzzy implication if*

(i) I is decreasing in the first variable and increasing in the second variable,
(ii) $I(0, 0) = I(1, 1) = 1$ and $I(1, 0) = 0$.

Let the set of all fuzzy implications be denoted by $\mathbb{I}$. For more on fuzzy implications, please see [1].

Among the generating methods of fuzzy implications from fuzzy implications proposed in [4], the following method gives not only new fuzzy implications but also a rich algebraic structure namely, monoid, on the set $\mathbb{I}$ of all fuzzy implications. In this connection we recall few results from [4].

Definition 2. *([4]) For $I, J \in \mathbb{I}$, define $I \circledast J$ as*

$$(I \circledast J)(x, y) = I(x, J(x, y)), \qquad x, y \in [0, 1].$$

Theorem 1. *([4]) $I \circledast J$ is an implication on $[0, 1]$, i.e., $I \circledast J \in \mathbb{I}$.*

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 563–568, 2013.

Theorem 2. *([4])* $(\mathbb{I}, \circledast)$ *forms a monoid, whose identity element is given by*

$$I_{\mathbf{D}}(x,y) = \begin{cases} 1, & \text{if } x = 0, \\ y, & \text{if } x > 0. \end{cases}$$

Definition 3. *([3], [6]) Let* $(S, \star), (T, \odot)$ *be two algebraic structures. A map* $f\colon (S,\star) \to (T,\odot)$ *is called a homomorphism if* $f(a \star b) = f(a) \odot f(b)$, *for all* $a, b \in S$.

Since homomorphisms play an important role in Algebra, the authors in [5] defined a particular function on $(\mathbb{I}, \circledast)$ by fixing a fuzzy implication in the following way.

Definition 4. *([5]) For any fixed* $K \in \mathbb{I}$, *define* $g_K\colon (\mathbb{I}, \circledast) \to (\mathbb{I}, \circledast)$ *by*

$$g_K(I) = I \circledast K, \qquad I \in \mathbb{I}.$$

In [5], it was proved that for every $K \in \mathbb{I}$, the map g_K is a lattice homomorphism. However g_K is not a semigroup homomorphism (s.g.h.) for every $K \in \mathbb{I}$(See Example 28 in [5]). Thus it has become necessary to characterise and represent $K \in \mathbb{I}$ such that g_K is an s.g.h. In the same work [5], we have investigated the class of trivial range implications and nontrivial range neutral implications K such that g_K is an s.g.h.

However, the class of nontrivial range non neutral implications K such that g_K is an s.g.h is not known. In this work, we give complete characterisation and representation results for this class and complete the study of such s.g.h g_K.

2 Semigroup Homomorphisms on $(\mathbb{I}, \circledast)$.

2.1 Trivial Range Implication K

The characterisation and representation of fuzzy implications K such that g_K, defined as in Definition 4, is an s.g.h was completely obtained in the case of $K(x,y) \in \{0,1\}$, for all $x, y \in [0,1]$, please see [5] for more details. We present below the main result of this case.

Theorem 3. *([5]) Let* $K \in \mathbb{I}$ *be such that the range of* K *is trivial. Then the following statements are equivalent:*

(i) g_K *is an* s.g.h.
(ii) $K = K^{\delta}$ *for some* $\delta \in]0,1]$ *where*

$$K^{\delta}(x,y) = \begin{cases} 1, & \text{if } x < 1 \text{ or } (x = 1 \text{ and } y \geq \delta)\ , \\ 0, & \text{otherwise}\ . \end{cases}$$

2.2 Nontrivial Range Implication K

In the case of nontrivial range, characterisation and representation of K such that g_K is an s.g.h. was done only for implications K satisfying $K(1, y) = y$ for all $y \in [0, 1]$. However, the class of nontrivial implications for which $K(1, y) \neq y$ for some $y \in (0, 1)$ has to be found out such that the function g_K is an s.g.h. Before doing this we investigate the range of K such that g_K is an s.g.h.

Lemma 1. *If the range of K is nontrivial and g_K is an s.g.h then the range of K is equal to $[0, 1]$.*

In order to obtain the representation of nontrivial K such that g_K is an s.g.h, we characterised the vertical section $K(1, \cdot)$ of K in [5]. For a quick reference we present this result in the following.

Proposition 1. *([5]) Let the range of $K \in \mathbb{I}$ be nontrivial and g_K be an s.g.h. Then there exist $\alpha, \beta \in [0, 1]$ such that the vertical section $K(1, .)$ has the following form:*

$$K(1,y) = \begin{cases} 0, & \text{if } \ y \in [0, \alpha) \ , \\ 0 \ \textbf{or} \ \alpha, & \text{if } \ y = \alpha \ , \\ y, & \text{if } \ y \in (\alpha, \beta) \ , \\ \beta \ \textbf{or} \ 1, & \text{if } \ y = \beta \ , \\ 1, & \text{if } \ y \in (\beta, 1] \ . \end{cases} \tag{1}$$

The following definition helps us in getting representation of K.

Definition 5. *Let $K \in \mathbb{I}$ and g_K is an s.g.h. Define two real numbers ϵ_0, ϵ_1 in the following way:*

$$\epsilon_0 = \sup\{t \in [0, 1] | K(1, t) = 0\},$$
$$\epsilon_1 = \inf\{t \in [0, 1] | K(1, t) = 1\} \ .$$

For every $K \in \mathbb{I}$, since $K(1, 0) = 0$ and $K(1, 1) = 1$, the real numbers ϵ_0, ϵ_1 in Definition 5 are well defined and exist in general. In the following we investigate the possible values of ϵ_0, ϵ_1 for $K \in \mathbb{I}$ such that range of K is nontrivial.

Lemma 2. $\epsilon_1 \neq 0$.

Proof. Let $\epsilon_1 = 0$. This implies that $K(1, y) = 1$ for all $y > 0$. It follows from the monotonicity of I in the first variable that $K(x, y) = 1$ for all $x \in [0, 1]$ and $y > 0$. So the range of K becomes $\{0, 1\}$, a contradiction to the fact the range of K is nontrivial. Thus $\epsilon_1 \neq 0$.

Lemma 3. *If $\epsilon_0 = 0$, then $\epsilon_1 = 1$. This implies that $K(1, y) = y$ for all $y \in [0, 1]$.*

Proof. Let $\epsilon_0 = 0$ and suppose that $\epsilon_1 < 1$. Then from Lemma 2, it follows that $0 < \epsilon_1 < 1$. So choose a $\delta > 0$ such that $0 < \epsilon_1 + \delta < 1$. Let $0 < y_1 < \epsilon_1$. This implies that $0 < K(1, y_1) = \alpha < 1$. Now, choose a $J \in \mathbb{I}$ such that $J(1, K(1, y_1)) = J(1, \alpha) = \epsilon_1 + \delta$. But $K(1, J(1, K(1, y_1))) = K(1, \epsilon_1 + \delta) = 1$, which contradicts g_K is an s.g.h. Thus $\epsilon_1 = 1$.

Lemma 4. *If $0 < \epsilon_0 < 1$, then $\epsilon_0 \neq \epsilon_1$.*

Proof. Let $0 < \epsilon_0 < 1$. Suppose that $\epsilon_0 = \epsilon_1$. Then $K(1, \cdot)$ will be of the form

$$K(1, y) = \begin{cases} 1, & \text{if } y \geq \epsilon_0 \\ 0, & \text{if } y < \epsilon_0 \end{cases} \tag{2}$$

This implies that $K(x, y) = 1$ for all $x \in [0, 1], y \geq \epsilon_0$. Now we prove that $K(x, y) = 1$ for all $x \in [0, 1[, y \in [0, \epsilon_0[$. On the contrary suppose that $\alpha = K(x_0, y_0) < 1$ for some $x_0 \in]0, 1[$, $y_0 \in [0, \epsilon_0[$. Since $0 < \epsilon_0 < 1$, choose a $\delta > 0$ such that $0 < \epsilon_0 + \delta < 1$. Now choose a $J \in \mathbb{I}$ such that $J(x_0, K(x_0, y_0)) = J(x_0, \alpha) = \epsilon_0 + \delta \neq 1$. Now, $K(x_0, J(x_0, K(x_0, y_0))) = K(x_0, J(x_0, \alpha)) = K(x_0, \epsilon_0 + \delta) = 1$ a contradiction to the fact that g_K is an s.g.h. Thus $K(x_0, y_0) = 1$ for all $x_0 \in [0, 1[$ and $y_0 \in [0, \epsilon_0[$ and $K(x, y) = 1$ for all $x < 1$. Finally from (2) it follows that the range of K is trivial, a contradicition. Thus $\epsilon_0 \neq \epsilon_1$.

Lemma 5. *If $\epsilon_0 > 0$, then $\epsilon_0 = 1$. This implies that $K(1, y) = 0$ for all $y > 0$.*

Proof. Let $\epsilon_0 > 0$. Suppose $\epsilon_0 < 1$, i.e., $0 < \epsilon_0 < 1$. Now Lemma 4 implies that $\epsilon_0 \neq \epsilon_1$. Now let $y_1 \in]\epsilon_0, \epsilon_1[$. Then $K(1, y_1) = y_1$ by (1) of Proposition 1. Choose a $J \in \mathbb{I}$ be such that $J(1, y_1) = \frac{\epsilon_0}{2}$. Then $(J \circledast K)(1, y_1) = J(1, K(1, y_1)) = J(1, y_1) = \frac{\epsilon_0}{2}$ and $(K \circledast J \circledast K)(1, y_1) = K(1, J(1, K(1, y_1))) = K(1, J(1, y_1)) = K(1, \frac{\epsilon_0}{2}) = 0$ a contradiction to g_K is an s.g.h. Thus $\epsilon_0 = 1$.

From Lemmas 3, 5 it follows that if K is a non trivial range implication such that g_K is an s.g.h then K is such that $K(1, y) = y$ for all $y \in [0, 1]$ or $K(1, y) = 0$ for all $y \neq 1$. Once again here we recall that the characterisation of nontrivial K such that g_K is an s.g.h was completely done in the case K satisfies $K(1, y) = y$ for all $y \in [0, 1]$. Here in the following we present main result of this issue. For more results, please see [5].

Definition 6. *([5]) For $\epsilon \in [0, 1[$ define*

$$K_\epsilon(x, y) = \begin{cases} 1, & \text{if } x \leq \epsilon, \\ y, & \text{if } x > \epsilon, \end{cases}$$

Note that $K_\epsilon \in \mathbb{I}$, for all $\epsilon \in [0, 1]$ and $\sup K_\epsilon = I_{\mathbf{WB}}$ where

$$I_{\mathbf{WB}}(x, y) = \begin{cases} 1, & \text{if } x < 1, \\ y, & \text{if } x = 1. \end{cases} \tag{3}$$

For notational convenience, we denote the set of all such K_ϵ implications by

$$\mathbb{K}_\epsilon^+ = \{I \in \mathbb{I} | I = K_\epsilon \text{ for some } \epsilon \in [0, 1[\} \cup I_{\mathbf{WB}}.$$

Theorem 4. *([5]) Let $K \in \mathbb{I}$ be satisfy $K(1, y) = y$ for all $y \in [0, 1]$. Then the following statements are equivalent:*

(i) g_K is an s.g.h.
(ii) $K \in \mathbb{K}_\epsilon^+$.

Now it remains to characterise the nontrivial range non neutral implications K such that g_K is an s.g.h. Now we will take up this in the following subsection.

2.3 Characterisation and Representation of K satisfying $K(1, y) = 0$ for all $y < 1$ such that g_k is an s.g.h.

Definition 7. *For* $\epsilon \in [0, 1[$ *define*

$$K^{\epsilon}(x, y) = \begin{cases} 1, & \text{if } x \leq \epsilon \ , \\ y, & \text{if } \epsilon < x < 1 \ , \\ 0, & \text{if } x = 1 \ \& \ y \neq 1 \ . \end{cases} \tag{4}$$

For notational convenience, we denote the set of all such K^{ϵ} implications by

$$\mathbb{K}^{\epsilon} = \{I \in \mathbb{I} | I = K^{\epsilon} \text{ for some } \epsilon \in [0, 1[\}.$$

Theorem 5. *Let* $K \in \mathbb{I}$ *be such that* $K(1, y) = 0$ *for all* $y \neq 1$. *Then* g_K *is an* s.g.h $\iff K \in \mathbb{K}^{\epsilon}$ *for some* $\epsilon \in [0, 1[$.

Proof. Let $K \in \mathbb{I}$ such that $K(1, y) = 0$ for all $y \neq 1$.
$(\Longrightarrow)$: Let g_K be an *s.g.h* for some $K \in \mathbb{I}$. Since the range of K is non-trival, from Lemma 1, it follows that the range of K is $[0, 1]$. Let $0 < \alpha < 1$ be chosen arbitrarily. Then there exist some $x_0 \in]0, 1[, y_0 \in [0, 1[$, such that $0 < K(x_0, y_0) = \alpha < 1$. We keep K fixed, vary J and investigate the equivalence $J \circledast K = K \circledast J \circledast K$.

When $J = I_{\mathbf{0}}$, we have

$$(J \circledast K)(x_0, y_0) = I_{\mathbf{0}}(x_0, K(x_0, y_0)) = I_{\mathbf{0}}(x_0, \alpha) = 0 \ ,$$
$$(K \circledast J \circledast K)(x_0, y_0) = K(x_0, I_{\mathbf{0}}(x_0, K(x_0, y_0))) = K(x_0, 0).$$

Since g_K is an *s.g.h.*, $K(x_0, 0) = 0$. Hence, if $K(x_0, y_0) = \alpha < 1$, then $K(x_0, 0) = 0$. Now, for any $J \in \mathbb{I}$, we have

$$(J \circledast K)(x_0, 0) = J(x_0, K(x_0, 0)) = J(x_0, 0)$$
$$\text{and } (K \circledast J \circledast K)(x_0, 0) = K(x_0, J(x_0, K(x_0, 0))) = K(x_0, J(x_0, 0)).$$

Now let us, once again, choose $J \in \mathbb{I}$ such that $J(x_0, 0) = y_0$. Then

$$y_0 = J(x_0, 0) = K(x_0, J(x_0, 0)) = K(x_0, y_0) = \alpha.$$

Since α is chosen arbitrarily, we have

$$K(x_0, y) = y, \qquad y \in [0, 1]. \tag{5}$$

Let $x^* = \inf\{x | K(x, y) = y, \text{ for all } \ y\} \geq 0$. Note that the infimum exists because x_0 satisfies (5).

Claim: $K(s, y) = 1$, for any $s \in [0, x^*[$ and for all $y \in [0, 1]$.

Proof of the claim: On the contrary, let us suppose that $1 > K(s, y_0) = y_1 > y_0$ for some y_0, y_1. Now,

$$J(s, K(s, y_0)) = J(s, y_1) \ ,$$
$$K(s, J(s, K(s, y_0))) = K(s, J(s, y_1)) \ .$$

Once again, choosing a $J \in \mathbb{I}$ such that $J(s, y_1) = y_0$, we get

$$J(s, y_1) = y_0 \text{ and } K(s, J(s, y_1)) = K(s, y_0) = y_1 \ ,$$
$$\implies J(s, K(s, y_0)) \neq K(s, J(s, K(s, y_0))),$$

i.e., g_K is not an *s.g.h.*, a contradiction. Thus $K(s, y) = 1$, for all $s \in [0, x^*[$.

Now the question is what value should one assign to $K(x^*, y)$. To allow for the possibility that $x^* = 0$ and since it is customary to assume left-continuity of fuzzy implications in the first variable, we let $K(x^*, y) = 1$. Note that letting $K(x^*, y) = y$ also gives a K such that g_K is a homomorphism.

From the above claim and (5) we see that every K is of the form (4) for some $\epsilon \in [0, 1[$.

($\Longrightarrow$): This follows easily.

3 Conclusions

The implications K for which the map g_K defined as in Definition 4 is an s.g.h were characterised and their representations were given in the case where K has trivial range. Further in the nontrivial range case the same was done for K satisfying $K(1, y) = y$ for all $y \in [0, 1]$. In this paper we showed that in the second case there are only two classes of nontrivial range $K \in \mathbb{I}$ for which g_K is an s.g.h. Also we characterised and found the representation of the nontrivial non neutral implications K for which g_K is an s.g.h thus completing the characterisations and representations of $K \in \mathbb{I}$ for which g_K is an s.g.h in all the cases.

Acknowledgements. The authors would like to acknowledge CSIR HRDG-INDIA (09/1001/(0008)/2011-EMR-I), DST -INDIA, (SR/FTP/MS-030/2010), respectively, for their financial support.

References

1. Baczyński, M., Jayaram, B.: Fuzzy implications. STUDFUZZ, vol. 231, pp. 1–35. Springer, Heidelberg (2008)
2. Fodor, J.C., Roubens, M.: Fuzzy preference modelling and multicriteria decision support. Kluwer, Dordrecht (1994)
3. Howie, J.M.: Fundamentals of semigroup theory. Oxford University Press (1995)
4. Vemuri, N.R., Jayaram, B.: Fuzzy implications: Novel generation process and the consequent algebras. In: Greco, S., Bouchon-Meunier, B., Coletti, G., Fedrizzi, M., Matarazzo, B., Yager, R.R. (eds.) IPMU 2012, Part II. CCIS, vol. 298, pp. 365–374. Springer, Heidelberg (2012)
5. Vemuri, N.R., Jayaram, B.: Homomorphisms on the monoid $(\mathbb{I}, \circledast)$. In: Proceedings of the 2013 International Conference on Fuzzy Systems(FUZZ-IEEE 2013), Hyderabad, July 7-10 (2013), doi:10.1109/FUZZ-IEEE.2013.6622436
6. Rotman, J.A.: An introduction to the Theory of Groups. Springer (1991)

Monotonicity of SISO Fuzzy Relational Inference Mechanism with Yager's Class of Fuzzy Implications

Sayantan Mandal and Balasubramaniam Jayaram

Department of Mathematics
Indian Institute of Technology Hyderabad,
Yeddumailaram-502205, India
{ma10p002,jbala}@iith.ac.in

Abstract. Monotonicity of a Fuzzy Relational Inference (FRI) mechanism has been studied in the literature with implicative form of a rule base where the implication comes from a residuated lattice structure. Here we show monotonicity of an FRI with Yager's class of fuzzy implications which do not form a residuated lattice structure on $[0, 1]$.

Keywords: monotone rule base, f-implications, fuzzy relational inference.

1 Introduction

Fuzzy Relational Inferences (FRI) occupy a central position in Approximate Reasoning using fuzzy logic. Two of the well-known Fuzzy Relational Inference mechanisms are the Compositional Rule of Inference (CRI) proposed by Zadeh [10], [4], and the Bandler-Kohout Subproduct (BKS) proposed by Pedrycz [6] based on the earlier work of Bandler and Kohout [2].

Monotonicity is one of the essential properties of an inference mechanism, unavailibility of which leads to an unreliable inference mechanism, see [3], [8]. Monotonicity of an FRI depends essentially on the operations employed in the FRI. The study of monotonicity in an FRI using implicative form of rules has largely been confined to operations that come from a residuated lattice. In this work, we investigate the monotonicity of an FRI where the underlying operations do not come from a residuated structure. Specifically, we study FRIs where the implicative rules are modeled based on the Yager's family of f-implications and present conditions underwhich monotonicity of the final output can be ensured.

2 Fuzzy Sets Fuzzy Rules and Fuzzy Logic Connectives

2.1 Fuzzy Sets

Definition 1. *If X is a non-empty set then $\mathcal{F}(X)$ is the fuzzy power set of X, i.e., $\mathcal{F}(X) = \{A | A : X \to [0, 1]\}$.*

Definition 2. *A fuzzy set A is said to be*

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 569–574, 2013.

- normal *if there exists an* $x \in X$ *such that* $A(x) = 1$,
- convex *if* X *is a linear space and for any* $\lambda \in [0,1]$, $x, y \in X$, $A(\lambda x + (1-\lambda)y) \geq \min\{A(x), A(y)\}$.

Definition 3. *For an* $A \in \mathcal{F}(X)$, *the* Support, Height, Ceiling *and* α*-cut for an* $\alpha \in (0,1]$ *are, respectively denoted as Supp A, Hgt A, Ceil A and* $[A]_\alpha$, *and are defined as:*

$$\begin{aligned} \textit{Supp } A &= \{x \in X | A(x) > 0\} \ , \\ \textit{Hgt } A &= \sup\{A(x) | x \in X\} \ , \\ \textit{Ceil } A &= \{x \in X | A(x) = \textit{Hgt } A\} \ , \\ [A]_\alpha &= \{x \in X | A(x) \geq \alpha\} \ . \end{aligned}$$

A is said to be bounded *if Supp A is a bounded set.*

Definition 4. *Let* $\{A_i\}_{i=1}^n \subseteq \mathcal{F}(X)$ *be an arbitrary collection of fuzzy sets of* X . $\{A_i\}_{i=1}^n$ *is said to form a* fuzzy partition *on* X *if*

$$X \subseteq \bigcup_{i=1}^{n} \textit{Supp } A_i \ .$$

In literature, a partition of X as defined above is also called a **complete** partition.

Definition 5. *A fuzzy partition* $\{A_i\}_{i=1}^n \subseteq \mathcal{F}(X)$ *is said to be* ***Ruspini Partition*** *if* $\sum_{i=1}^{n} A_i(x) = 1$ *for every* $x \in X$.

Definition 6. *Let* $C \in \mathcal{F}(X)$ *and* $I \in \mathcal{I}$ *be any fuzzy implication.*

- *We say that a* $C' \in \mathcal{F}(X)$ *is the* modification or modified fuzzy set of C by I at a given $\alpha \in [0,1]$ *if*

$$C'(x) = I(\alpha, C(x)), \ x \in X \ . \tag{1}$$

- *Given two convex fuzzy sets* $C_1, C_2 \in \mathcal{F}(X)$, *we say that* $C_1 \prec C_2$ *if for any* $\alpha \in (0,1]$ *it holds that* $\inf[C_1]_\alpha \leq \inf[C_2]_\alpha$ *and* $\sup[C_1]_\alpha \leq \sup[C_2]_\alpha$.

The following results show that modification by an $I \in \mathcal{I}$ preserves convexity. We omit the proof of the following results due to space constraints.

Proposition 1. *For a convex fuzzy set* C, *a fuzzy implication* I *and any* $\alpha \in [0,1]$, *modified fuzzy set* $C' = I(\alpha, C)$ *is also convex.*

2.2 Defuzzification

Often there is a need to convert a fuzzy set to a crisp value, a process which is called *Defuzzification*. This process of defuzzification can be seen as a mapping $g : \mathcal{F}(X) \longrightarrow X$. There are many types of defuzzification techniques available in the literature, see [7] for a good overview. In this work, we use the following defuzzifier extensively.

Example 1. For an $A \in \mathcal{F}(X)$, the *Mean of Maxima* (MeOM) defuzzifier gives as output the mean of all those values in X with the highest membership value, which can be mathematically expressed as

$$\text{MeOM}(A) = \frac{\sum_{x \in \text{Ceil}A} x}{|\text{Ceil}(A)|} \ \text{ or } \ \frac{\int_{x \in \text{Ceil}A} dx}{|\text{Ceil}(A)|}. \tag{2}$$

depending on whether Ceil A is finite or not.

2.3 Fuzzy Rules

Given two non-empty classical sets $X, Y \subseteq \mathbb{R}$, a Single-Input Single-Output (SISO) fuzzy IF-THEN rulebase consists of rules of the form:

$$\textbf{IF } \tilde{x} \text{ is } A_i \textbf{ THEN } \tilde{y} \text{ is } B_i \ , \tag{3}$$

where $\tilde{x}$,$\tilde{y}$ are the linguistic variables and $A_i, B_i, i = 1, 2, \ldots n$ are the linguistic values taken by the linguistic variables. These linguistic values are represented by fuzzy sets in their corresponding domains, i.e., $A_i \in \mathcal{F}(X), B_i \in \mathcal{F}(Y)$.

Definition 7 ([8]). *A fuzzy rule base* (3) *is called* monotone *if for any two rules :*

$$\textbf{\textit{IF }} \tilde{x} \textit{ is } A_i \textbf{\textit{ THEN }} \tilde{y} \textit{ is } B_i$$
$$\textbf{\textit{IF }} \tilde{x} \textit{ is } A_j \textbf{\textit{ THEN }} \tilde{y} \textit{ is } B_j \ ,$$

such that $A_i \prec A_j$, the inequality $B_i \prec B_j$ holds.

2.4 Fuzzy Logic Connectives

Definition 8 ([1], Definition 1.1.1). *A function $I\colon [0,1]^2 \longrightarrow [0,1]$ is called a* fuzzy implication *if it is decreasing in the first variable, increasing in the second variable and $I(0,0) = 1, I(1,1) = 1, I(1,0) = 0$.*

Yager [9] introduced a family of fuzzy implications based on strictly monotonic functions on $[0,1]$.

Definition 9 ([1], Definition 3.1.1). *Let $f\colon [0,1] \to [0,\infty]$ be a strictly decreasing and continuous function with $f(1) = 0$. The function $I_f\colon [0,1]^2 \to [0,1]$ defined by*

$$I_f(x,y) = f^{-1}\left(x \cdot f(y)\right) \ , \qquad x, y \in [0,1] \ , \tag{4}$$

*with the understanding $0 \cdot \infty = 0$, is a fuzzy implication and called an f-*implication.

The function f itself is called an f-generator of the I_f generated as in (4). We will often write $\longrightarrow_f$ instead of I_f. We can divide the whole class of f-implication into two classes depending on whether $f(0) = \infty$ or $f(0) < \infty$. In this article we only consider the class of f-implications for which $f(0) = \infty$, in which case, it is easy to verify that $x \longrightarrow_f 0 = 0$ for any $x \in (0,1]$.

3 Fuzzy Relational Inference Mechanism

Given a rule base of the form (3) and an input " $\tilde{x}$ is A' ", the main objective of a fuzzy inference mechanism is to find a *meaningful* B' such that " $\tilde{y}$ is B' ". While many types of fuzzy inference mechanisms have been proposed in the literature we restrict this study only to Fuzzy Relation based inference mechanisms.

The inference mechanism in a fuzzy relational inference (FRI) can be expressed as follows:

$$B' = f_R^{@}(A') = A'@R \ , \tag{FRI-R}$$

where $A' \in \mathcal{F}(X)$ is the input, the relation $R : X \times Y \to [0,1]$ i.e, $R \in \mathcal{F}(X \times Y)$ represents or models the rule base, B' is the obtained output and $@$ is called the *composition operator*, which is a mapping $@\colon \mathcal{F}(X) \times \mathcal{F}(X \times Y) \to \mathcal{F}(Y)$.

3.1 FRI with Singleton Input : $f_R^{@} \equiv f_R$

Often one needs to deal with crisp inputs $x_0 \in X$. In such a case the given input is suitably *fuzzified*, i.e., a fuzzy set $A' \in \mathcal{F}(X)$ is suitably constructed from x_0. Commonly, the following singleton fuzzifier is employed:

$$A'(x) = \begin{cases} 1, & x = x_0 \ , \\ 0, & x \neq x_0 \ . \end{cases}$$

With the above input A', the FRI mechanism (FRI-R) reduces to

$$B'(y) = R(x_0, y), \qquad y \in Y. \tag{FRI-R-Singleton}$$

Thus in the case of a singleton input, the output is fully dependent on the model of the rule base R. In other words, $f_R^{@} \equiv f_R$ and hence the composition @ does not play any role.

We denote an FRI with singleton input as a quadruple $\mathbb{F} = \Big(\{A_i, B_i\}, R, g\Big)$, where A_i's and B_i's correspond to the input and output fuzzy partitions on X and Y, respectively, R is the fuzzy relation modeling the rule base and g is the defuzzifier used to obtain a crisp output from the obtained B'.

3.2 FRI with Yager's Class of Fuzzy Implications

In this work, we consider the Relational inference mechanism, where the fuzzy implication is f-implication, one of the Yager's classes of fuzzy implications. Essentially, we interpret the $\longrightarrow$ in (FRI-R-Singleton) as an f-implication with $f(0) = \infty$ and consider the following relation R to represent the rule base:

$$\hat{R}_f(x, y) = \bigwedge_{i=1}^{n} (A_i(x) \longrightarrow_f B_i(y)), \qquad x \in X, y \in Y \ . \tag{Imp-$\hat{R}_f$}$$

Recently, we have studied the suitability of the Bandler-Kohout Subproduct FRI with (Imp-$\hat{R}_f$) representing the implicative form of the rule base in [5]. In this work, we consider (Imp-$\hat{R}_f$) and discuss the issue of monotonicity in this context.

4 Monotonicity of FRI with f- Implication(SISO Case)

Theorem 1. *Let us consider an FRI* $\mathbb{F} = \Big(\{A_i, B_i\}, R, g\Big)$ *with* $R = \hat{R}_f$ *with a monotone rule base of the form* (3) *such that* $\{A_i\}_{i=1}^n$, $\{B_i\}_{i=1}^n$ *are normal, convex, strict on both sides of the ceiling and form a Ruspini Partition on* X *and* Y *respectively. Then* $\mathbb{F}$ *with* MeOM *defuzzification as in* (2) *is monotonic.*

Proof. While the proof is valid for any fuzzy sets which are normal, convex and strict on both sides of ceiling, for better readability we prove this result for triangular fuzzy sets only. Fuzzy relational inference mechanism (FRI-R-Singleton) with $R = \hat{R}_f$ is of the form,

$$B'(y) = \hat{R}_f(x_0, y), \quad y \in Y. \tag{5}$$

Let the convex fuzzy sets $\{A_i\}_{i=1}^n, \{B_i\}_{i=1}^n$ be such that Supp $A_i = [x_{i-1}, x_{i+1}]$, Supp $B_i = [y_{i-1}, y_{i+1}]$ and $A_i(x_i) = 1$, $B_i(y_i) = 1$. Let $x', x'' \in X$ be the two given inputs such that $x' \leq x''$. We consider the following two cases and show that the obtained outputs $y', y'' \in Y$ also are similarly ordered, i.e., $y' \leq y''$.

Case-1: $x', x'' \in [x_i, x_{i+1}]$, for $i = 1, 2, \ldots, n-1$.

Since $\{A_i\}_{i=1}^n$ form Ruspini partition $A_j(x') = 0 \ \ \forall j \neq i, i+1$. From (5),

$$B'(y) = [A_i(x') \longrightarrow_f B_i(y)] \wedge [A_{i+1}(x') \longrightarrow_f B_{i+1}(y)].$$

Let $A_i(x') = \alpha_i'$ and $A_{i+1}(x') = \alpha_{i+1}'$. Since A_i's form a Ruspini partition, $\alpha_i' + \alpha_{i+1}' = 1$. Similarly, with $A_i(x'') = \alpha_i''$ and $A_{i+1}(x'') = \alpha_{i+1}''$, we have $\alpha_i'' + \alpha_{i+1}'' = 1$. Thus,

$$\begin{aligned} B'(y) =& [\alpha_i' \longrightarrow_f B_i(y)] \wedge [\alpha_{i+1}' \longrightarrow_f B_{i+1}(y)] = B_i'(y) \wedge B_{i+1}'(y) \ , \\ B''(y) =& [\alpha_i'' \longrightarrow_f B_i(y)] \wedge [\alpha_{i+1}'' \longrightarrow_f B_{i+1}(y)] = B_i''(y) \wedge B_{i+1}''(y) \ . \end{aligned}$$

Now since A_i is decreasing in $[x_i, x_{i+1}]$ and A_{i+1} is increasing in $[x_i, x_{i+1}]$,

$$x' \leq x'' \text{ and } x', x'' \in [x_i, x_{i+1}] \Longrightarrow A_i(x') \geq A_i(x'') \text{ and } A_{i+1}(x') \leq A_{i+1}(x'').$$

i.e., $\alpha_i' \geq \alpha_i''$ and $\alpha_{i+1}' \leq \alpha_{i+1}''$. Clearly, since B_i, B_{i+1} are convex and normal, B_i', B_{i+1}' are also convex and normal. So $B' = B_i' \cap B_{i+1}'$ and $B'' = B_i'' \cap B_{i+1}''$ are also convex. Let $y' = \text{MeOM}(B')$ and $y'' = \text{MeOM}(B'')$.

Claim: If $x' \in [x_i, x_{i+1}]$, then $y' \in [y_i, y_{i+1}]$ for $i = 1, 2, \ldots, n-1$.

Now since $x \longrightarrow_f 0 = 0$ for any $x \in (0, 1]$, we have that the supports of both the modified fuzzy sets $B_i' = \alpha_i \longrightarrow_f B_i, B_{i+1}' = \alpha_{i+1} \longrightarrow_f B_{i+1}$ are the same as those of B_i, B_{i+1}, i.e., Supp B_i'=Supp B_i and Supp B_{i+1}' = Supp B_{i+1}. Hence,

$$\begin{aligned} \text{Supp } B' &= \text{Supp } B_i' \cap \text{Supp } B_{i+1}' = \text{Supp } B_i \cap \text{Supp } B_{i+1} \\ &= \text{Supp } (B_i \cap B_{i+1}) = [y_i, y_{i+1}] \ . \end{aligned} \tag{6}$$

Since (6) holds we have,

$$y' = g(x') = \text{MeOM}(B'(y)) \in \text{Supp } (B_i \cap B_{i+1}) = [y_i, y_{i+1}].$$

So, $y', y'' \in [y_i, y_{i+1}]$ and we obtain the following inequalities, for any $y \in [y_i, y_{i+1}]$:

$$\begin{aligned} \alpha_i' \geq \alpha_i'' \Longrightarrow & \alpha_i' \longrightarrow_f B_i(y) \leq \alpha_i'' \longrightarrow_f B_i(y), \\ \Longrightarrow & B_i'(y) \leq B_i''(y) \ . \\ \text{Similarly, } \alpha_{i+1}' \leq \alpha_{i+1}'' \Longrightarrow & \alpha_{i+1}' \longrightarrow_f B_{i+1}(y) \geq \alpha_{i+1}'' \longrightarrow_f B_{i+1}(y), \\ \Longrightarrow & B_{i+1}'(y) \geq B_{i+1}''(y) \ . \end{aligned}$$

Since B_i's are triangular fuzzy sets, $y', y'' \in [y_i, y_{i+1}]$ are the points which satisfy

$$B_i'(y') = B_{i+1}'(y') \Longrightarrow \alpha_i' \longrightarrow_f B_i(y') = \alpha_{i+1}' \longrightarrow_f B_{i+1}(y') \ , \tag{7}$$

$$B_i''(y'') = B_{i+1}''(y'') \Longrightarrow \alpha_i'' \longrightarrow_f B_i(y'') = \alpha_{i+1}'' \longrightarrow_f B_{i+1}(y'') \ . \tag{8}$$

Now, to prove monotonicity, we need to show that $y' \leq y''$.

If possible, let us assume to the contrary that $y' > y''$. Since B_i and B_{i+1} are, respectively, strictly decreasing and strictly increasing on $[y_i, y_{i+1}]$,

$$y' > y'' \ \text{ and } \ y', y'' \in [y_i, y_{i+1}] \Longrightarrow B_i(y') < B_i(y'') \ \text{ and } \ B_{i+1}(y') > B_{i+1}(y'') \ . \tag{9}$$

This leads to the following inequalities:

$$\begin{aligned}
\alpha'_{i+1} \longrightarrow_f B_{i+1}(y') &> \alpha'_{i+1} \longrightarrow_f B_{i+1}(y'') && \text{(Since } \longrightarrow_f \text{ is strict and using (9))} \\
&\geq \alpha''_{i+1} \longrightarrow_f B_{i+1}(y'') && \text{(Since } \alpha'_{i+1} \leq \alpha''_{i+1}\text{)} \\
&= \alpha''_i \longrightarrow_f B_i(y'') && \text{(Using (8))} \\
&> \alpha''_i \longrightarrow_f B_i(y') && \text{(Using (9))} \\
&\geq \alpha'_i \longrightarrow_f B_i(y') \ , && \text{(Since } \alpha'_i \geq \alpha''_i\text{)}
\end{aligned}$$

i.e., $\alpha'_{i+1} \longrightarrow_f B_{i+1}(y') > \alpha'_i \longrightarrow_f B_i(y')$, a contradiction to (7). So we have $x' \leq x'' \Longrightarrow y' \leq y''$.

Case-2: Let $x' \in [x_i, x_{i+1}]$ and $x'' \in [x_{i+p}, x_{i+p+1}], p \geq 1$. By the claim above, we have $y' \in [y_i, y_{i+1}]$ and $y'' \in [y_{i+p}, y_{i+p+1}], p \geq 1$ and hence $y' \leq y''$. □

5 Concluding Remarks

FRIs are one of the most important inference mechanisms to be studied. However, all of the known works on FRIs employ operations that come from a residuated lattice structure. In this work, we have shown that one of the most desirable properties, viz., *monotonicity*, is available to us with the operations coming from a non-residuated structure.

References

1. Baczyński, M., Jayaram, B.: Fuzzy Implications. STUDFUZZ, vol. 231, pp. 1–35. Springer, Heidelberg (2008)
2. Bandler, W., Kohout, L.J.: Semantics of implication operators and fuzzy relational products. Internat. J. Man-Mach. Stud. 12, 89–116 (1980)
3. Broekhoven, E.V., Baets, B.D.: Only Smooth Rule Bases Can Generate Monotone Mamdani-Assilian Models Under Center-of-Gravity Defuzzification. IEEE Transactions on Fuzzy Systems 17, 1157–1174 (2009)
4. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic contoller. Int. J. Man-Machine Studies 7, 1–13 (1975)
5. Mandal, S., Jayaram, B.: Bandler-Kohout Subproduct with Yagers classes of Fuzzy Implications. IEEE Trans. on Fuzzy Syst. (accepted 2013)
6. Pedrycz, W.: Application of fuzzy relational equations for methods of reasoning in presence of fuzzy data. Fuzzy Sets and Syst. 16, 163–175 (1985)
7. Roychowdhury, S., Pedrycz, W.: A survey of defuzzification strategies. International Journal of Intelligent Systems 16, 679–695 (2001)
8. Štěpnička, M., De Baets, B.: Monotonicity of implicative fuzzy models. In: IEEE International Conference on Fuzzy Systems, pp. 1–7 (2010)
9. Yager, R.R.: On some new classes of implication operators and their role in approximate reasoning. Inform. Sci. 167, 193–216 (2004)
10. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. on Syst. Man and Cyber. SMC-3(1), 28–44 (1973)

Small World Particle Swarm Optimizer for Global Optimization Problems

Megha Vora and T.T. Mirnalinee

Department of Computer Science and Engineering
S.S.N College of Engineering, Anna University, Chennai, India
meghavora25@gmail.com, mirnalineett@ssn.edu.in

Abstract. Particle swarm is a stochastic optimization paradigm inspired by the concepts of social psychology and artificial intelligence. Interrelationship between individuals in a swarm is defined by the population topology, which can be depicted as a network model. Regular networks are highly clustered but the characteristic path length grows linearly with the increase in number of vertices. On the contrary, random networks are not highly clustered but they have small characteristic path length. Small world network have a distinctive combination of regular and random networks *i.e.*, highly clustered and small characteristic path length. This paper takes forward the concept of incorporating small world theory in the Particle Swarm Optimization (PSO) framework. Efficiency of the proposed methodology is tested by applying it on twelve standard benchmark functions. Results obtained are compared with other PSO variants. Comparative study demonstrates the effectiveness of the proposed approach.

1 Introduction

Particle Swarm Optimization (PSO) is the global optimization technique originally proposed by Kennedy and Eberhart [1]. PSO is inspired by the paradigm of birds flocking. PSO consists of a swarm (population) of particles. Each particle flies through the multi-dimensional search space with a velocity. Velocity of a particle is updated using three terms i)previous velocity- provides necessary momentum to the particle, ii)social term- indicates how the particle is stochastically drawn towards the global best (*gbest*) position found so far by the swarm, and iii)cognitive term- reflects the personal thinking (*pbest*) of the particle. PSO has received a huge attention and popularity in the optimization community due to its algorithmic simplicity and effectiveness for solving global optimization problem. Moreover, it is computationally inexpensive in terms of both memory requirements and CPU speed. This derivative-free method is particularly suited to continuous variable problems. Even though PSO is a good and fast search algorithm, it suffers from premature convergence, especially in complex multi-peak-search problems. To overcome this problem in [2] dynamic network model is suggested; search begins with *pbest* and gradually incremented until it becomes *gbest*. In [3] Small World Particle Swarm Optimization (SWPSO) was

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 575–580, 2013.

proposed where *gbest* term in velocity updation was replaced by the small world best (*SWNbest*). Li *et al.*, [4] proposed Self Learning Particle Swarm Optimizer (SLPSO) where each particle has a set of four strategies to cope with different situations. A particle can choose an appropriate strategy at any instance according to the property of its local search space.

The population topology has a significant effect on the performance of PSO. Topology determines the way particles communicate and share information within the swarm . In [5] and [6] impact of various network topologies on particle swarm performance was studied. It was reported that path length presents a compromise between exploration and exploitation. If path length is too small, it implies that information spreads too fast, so there is a higher probability of premature convergence. On the contrary, if it is large then information takes a long time to travel through the graph. Thus, the population is more resilient and not so eager to exploit earlier on. Even a partial degree clustering (i.e., percentage of vertex's neighbours that are neighbours to one another) helps to disseminate information in the network. We therefore need to have a combination of highly clustered and small characteristic path length network. Small world network [7] fits to this need of ours.

Small world concept was originally proposed by Milgram in [8] which was later modeled by Watts-Strogatz [9]. Jon Kleinberg [7] further generalized Watts-Strogatz model and showed that there is a decentralized algorithm capable of finding short paths with high probability.

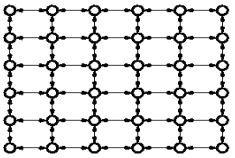


Fig. 1. Swarm of particles 6x6

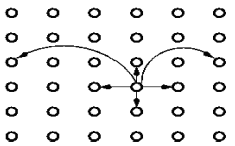


Fig. 2. Small world with 2 random particles

2 Proposed Algorithm - SWPSO-I

In [3], small world network topology proposed by Jon Kleinberg [7] on Von Neumann network model was considered. Fig. 1 shows Von Neumann network with swarm of size 6x6 and each particle has up, down, left and right particles in its neighborhood. Now, small world network is formed by adding few random particles (in this case two) and retaining four immediate Von Neumann neighbors of a current particle. These random particles are treated as additional neighbors of current particle (see Fig. 2). The intuitive advantage of these additional random neighbors is the way small world is formed within the current swarm.

In this paper, SWPSO-I algorithm is being proposed which takes forward the above mentioned concept of SWPSO. Different from the work in [3] where inertia weight parameter was set constant, in this paper linearly decreasing inertia weight has been used. This leads to a more balanced exploration-exploitation

trade-off [10]. Moreover, in this paper particle's velocity and position is updated at individual level rather than at population level to make it adaptive. Algorithmic detail of the SWPSO-I is given below.

1 Initialize the position $\overrightarrow{X}_k$ and velocity $\overrightarrow{V}_k$ of each k^{th} particle in the swarm where k varies from $(1, 2, ..., NP)$; NP is the size of swarm
2 Initialize fitness counter $fes = 0$. Define maximum number of fitness evaluation T_FEs
3 Compute the fitness F_k for each k^{th} particle in the swarm
4 Perform steps 5 to 13 until maximum number of fitness evaluation is not exceeded
5 For each of the k^{th} particle perform steps 6 to 13
6 Compare $Pbest_k$ of the k^{th} particle with its current fitness value F_k. If the current fitness value is better, then assign the current fitness value to $pbest_k$ and assign the current position to $\overrightarrow{X}_{Pbest_k}$
7 Compute the small world best fitness $SWNbest_k$ by selecting particle with best fitness (here minimum) in the small world neighbourhood of the k^{th} particle
8 Update the previous small world best fitness with the current value of $SWNbest_k$ if it is minimum. Also, update previous position of the best small world particle with the current one $\overrightarrow{X}_{SWNbest_k}$
9 Update velocity of the particle using Eq. 1

$$\overrightarrow{V}_k^{'} = \omega V_k + C_1 R_1(\overrightarrow{X}_{Pbest_k} - \overrightarrow{X}_k) + C_2 R_2(\overrightarrow{X}_{SWNbest_k} - \overrightarrow{X}_k) \quad (1)$$

where, ω is inertia weight as defined in Eq. 2; lb and ub represents lower and upper bound respectively; C_1 and C_2 are acceleration constant; R_1 and R_2 are random numbers uniformly generated from the interval $[0, 1]$

$$\omega(fes) = ub - (lb + 0.1 * fes)/T\_FEs \quad (2)$$

10 If $\overrightarrow{V}_k^{'} < -V_{max}$ then $\overrightarrow{V}_k^{'} = -V_{max}$ else if $\overrightarrow{V}_k^{'} > V_{max}$ then $\overrightarrow{V}_k^{'} = V_{max}$; V_{max} is set to the half of the search range
11 Update the position of the particle using Eq. 3

$$\overrightarrow{X}_k^{'} = \overrightarrow{X}_k + \overrightarrow{V}_k^{'} \quad (3)$$

12 Compute the fitness of the updated particle
13 Increment the fitness counter $fes++$

3 Experimental Results

3.1 Benchmark Functions Used

For evaluating the performance of SWPSO-I a test set of twelve functions is used. Details of the functions are given in Table 1. The test function set includes the traditional functions and traditional functions with noise, shift and rotation as described in [11]. Detail description of the Shifting, rotation and noise addition method for the test functions $f_8 - f_{12}$ can be found in [11].

Table 1. Test functions, where f_{min} is the minimum value of a function and S ϵ R_n

Name	Test Function	Search Range S	f_{min}
Sphere	$f_1(x) = \sum_{i=1}^{n} x_i^2$	[-100, 100]	0
Rastrigin	$f_2(x) = \sum_{i=1}^{n} (x_i^2 - 10\cos(2\Pi x_i) + 10)$	[-5.12, 5.12]	0
Griewank	$f_3(x) = \frac{1}{4000}\sum_{i=1}^{n}(x_i - 100)^2 - \prod_{i=1}^{n}\cos\frac{x_i - 100}{\sqrt{i+1}}$	[-600, 600]	0
Ackley	$f_4(x) = -20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) -$ $\exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)) + 20 + e$	[-32 32]	0
Rosenbrock	$f_5(x) = \sum_{i=1}^{n} 100(x_{i+1}^2 - x_i)^2 + (x_i - 1)^2$	[-2.048, 2.048]	0
Schwefel-1-2	$f_6(x) = \sum_{i=1}^{n}(\sum_{j=1}^{i} x_j^2)$	[-100, 100]	0
Schwefel-2-22	$f_7(x) = \sum_{i=1}^{n} \lvert x_i\rvert + \prod_{i=1}^{n} \lvert x_i\rvert$	[-10, 10]	0
S-Sphere	$f_8(x) = \text{CEC05}(F_1)$	[-100, 100]	-450
S-Rosenbrock	$f_9(x) = \text{CEC05}(F_6)$	[-100, 100]	390
S-schwefel-1-2	$f_{10}(x) = \text{CEC05}(F_2)$	[-100, 100]	-450
S-schwefel-1-2-Noisy	$f_{11}(x) = \text{CEC05}(F_4)$	[-100, 100]	-450
RS-Ackely	$f_{12}(x) = \text{CEC05}(F_8)$	[-32, 32]	-140

3.2 Simulation Strategy

To fairly compare the results obtained by SWPSO-I with SWPSO [3] and already published results of SLPSO [4], simulation was done considering a swarm of size of 20 (4x5). Swarm was initialized in the search space using symmetric initialization strategy (*i.e.*, uniformly distributed over the search space S). Functions were tested in 30-dimension (30D) with 30 independent runs. Each run of the algorithm is terminated when the number of function evaluations (T_FEs) exceed 1e+05. Acceleration coefficient parameter C_1 and C_2 is set to 1.494. These values were chosen to ensure good convergence [12]. Inertia weight parameter ω is varied linearly from 0.9 to 0.4 according to the Eq. 2 for both SWPSO-I and SLPSO, while it is set to 0.72 for SWPSO. The larger step size at the beginning offers a better exploration and the smaller step size towards the end offers better exploitation, thus decreasing chances of missing a global optima. For SWPSO and SWPSO-I two small world randomized particles were chosen. Experiments with different number (1, 2 and 3) of randomized particles showed that using two randomized particles convergence to optimal solution was faster.

3.3 Results on Benchmark Functions

Table 2 shows the mean and standard deviation (within parentheses) for 30 runs for three algorithms. Results in bold face indicate the best ones. Table 3 shows the result of two tailed unpaired t-tests [13] with 58 degree of freedom between the best algorithm and the second best in each case.

It can be observed from Table 2 that SWPSO-I outperforms SWPSO in terms of mean and standard deviation. Only in one case (f_7) SWPSO-I mean value is larger, but Table 3 shows that this difference is not statistically significant. For the function f_{12} both SWPSO-I and SWPSO have same mean value but low standard deviation value of SWPSO-I shows that SWPSO-I is more stable than SWPSO. Thus in terms of stability and accuracy SWPSO-I outperforms

Table 2. Comparison results of mean and standard deviation (within parentheses) in 30 Dimensions over 30 runs

f	f_1	f_2	f_3	f_4
SWPSO-I	**9.6e-71(5.2e-70)**	70.64(15.13)	**6.56e-3(8.91e-3)**	5.35e-1(7.63e-1)
SWPSO	2.94e-61(1.04e-60)	69.52(20.59)	9.3e-3(1.08e-2)	1.452(1.086)
SLPSO	2.78e-50(8.09e-49)	**0(0)**	0.227(0.144)	**3.47e-14(4.45e-14)**
f	f_5	f_6	f_7	f_8
SWPSO-I	6.2(**5.3**)	**5.11e-4(1.01e-3)**	1.75e-27(9.48e-27)	3.67e-13(7.09e-13)
SWPSO	10.68(12.75)	1.66e+2(9.12e+2)	**1.51e-29(8.29e-29)**	1.79e+2(9.85e+2)
SLPSO	**2.06** (12.3)	7.93e-3(0.233)	1.35e-26(3.88e-25)	**1.00e-13(1.74e-13**)
f	f_9	f_{10}	f_{11}	f_{12}
SWPSO-I	23.6(**47.05**)	2.06e-3(7.69e-3)	**1.46e+3(9.73e+2)**	20.9(**6.82e-2**)
SWPSO	1.15e+10(1.59e+8)	8.40(3.24e+1)	4.32e+3(4.95e+3)	20.9(0.0860)
SLPSO	**10.7**(140)	**1.4e-6(1.57e-6)**	1.16e+4(2.99e+4)	**20.4**(0.846)

SWPSO. Table 2 shows that for seven cases SWPSO-I gives higher mean (i,e, worse) than the mean of SLPSO algorithm; Table 3 shows that this difference is statistically not significant for five cases. Moreover for f_{12} function difference in mean values of SWPSO-I and SLPSO is not significant but from standard deviation point of view SWPSO-I is giving more stable result. Similar is the case for function f_9 and f_5. Thus SWPSO-I is more stable compared to SLPSO.

Table 3. Results of unpaired t-tests on the data of Table 2

Function	Std. Error	t	Two tailed p	Significance
f1	0	1.548	0.1270	Not significant
f2	3.759	18.493	<0.0001	Extremely significant
f3	0.003	1.071	0.2882	Not significant
f4	0.026	8.368	<0.0001	Extremely significant
f5	2.445	1.693	0.0958	Not quite significant
f6	0.043	0.174	0.8622	Not significant
f7	0.000	1.002	0.3203	Not significant
f8	0.0000	2.003	0.0498	Not significant
f9	26.965	0.478	0.6342	Not significant
f10	0.001	1.466	0.1480	Not significant
f11	922.537	3.097	0.0030	Very significant
f12	0.155	3.226	0.0021	Very significant

Table 4. Summarized t-test results

SWPSO-I Vs.	Significantly better	Significantly worse	Statistically equivalent
SWPSO	3	0	9
SLPSO	2	2	8

Table 4 summarizes t-test results. It is apparent from the table that performance of SWPSO-I is better than SWPSO and comparable with that of SLPSO.

4 Conclusions

In this paper, we have proposed SWPSO-I algorithm which takes forward the concept of SWPSO. To evaluate the performance of the proposed algorithm we have tested it on 12 benchmark functions. Mean, standard deviation and t-test are the performance metrics used for evaluation. Comparative study of the proposed algorithm with SWPSO and SLPSO justify the efficiency of the proposed SWPSO-I. Applying SWPSO-I for real world optimization problems and molding it for data clustering and image segmentation will be the future scope of study.

References

1. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
2. Suganthan, P.N.: Particle swarm optimizer with neighborhood operator. In: Congress on Evolutionary Computation, pp. 1958–1962 (1999)
3. Saxena, A.K., Vora, M.: Novel approach for the use of small world theory in particle swarm optimization. In: 16th International Conference on Advanced Computing and Communications, pp. 363–366. IEEE (2008)
4. Li, C., Yang, S., Nguyen, T.: A self-learning particle swarm optimizer for global optimization problems. IEEE Transactions on Systems, Man, and Cybernetics, Part B 42(3), 627–646 (2012)
5. Mendes, R.: Population Topologies and Their Influence in Particle Swarm Performance. PhD thesis, University of Minho (2004)
6. Kennedy, J.: Small worlds and mega-minds: Effects of neighborhood topology on particle swarm performance. In: IEEE CEC, pp. 1931–1938 (1999)
7. Kleinberg, J.: The small-world phenomenon: An algorithmic perspective. Technical report, Cornell University Ithaca, NY, USA (1999)
8. Milgram, S.: The small world problem. Psychology Today 2, 60–67 (1967)
9. Watts, D., Strogatz, S.: Collective dynamics of small-world networks. Nature 393, 440–442 (1998)
10. Eberhart, R.C., Shi, Y., Kennedy, J.: Swarm Intelligence. Morgan Kaufmann (2001)
11. Suganthan, P.N., Hansen, N., Liang, J.J., Deb, K., Chen, Y.P., Auger, A., Tiwari, S.: Problem definitions and evaluation criteria for the cec 2005 special session on real-parameter optimization. Technical report (2005)
12. van den Bergh, F.: An Analysis of Particle Swarm Optimizer. PhD thesis, Department of Computer Science, University of Petoria, South Africa (2002)
13. Seltman, H.J.: Experimental Design and analysis. Carnegie Mellon University (2012)

Ordered Solution Generation for Implicit AND/OR Search Spaces

Priyankar Ghosh, Partha Pratim Chakrabarti, and Pallab Dasgupta

Indian Institute of Technology Kharagpur, West Bengal, India 721302
{priyankar,ppchak,pallab}@cse.iitkgp.ernet.in

Abstract. In this paper we address the problem of generating alternative solutions for implicit AND/OR DAGs which are described using a start state and a set of transformation rules. Unlike the explicit representation of AND/OR graph, for implicitly specified AND/OR graphs, the transformation rules are used to dynamically make a portion of an AND/OR graph explicit as per requirement. We propose an algorithm, ASGAO*, for generating alternative solutions for implicitly specified AND/OR graphs. Experimental results in several domains show that, on an average, our proposed algorithm performs hundred times better with respect to running time and fifty times better with respect to space than the existing algorithm. In view of the renewed research interest towards applying AND/OR graphs in domains like graphical model, constraint satisfaction, and service composition, it is not hard to see useful applications of ASGAO*.

1 Introduction

Search methods in artificial intelligence are being increasingly used to solve a wide variety of optimization problems from diverse domains like scheduling, routing, constraint satisfaction, CAD for VLSI, etc. In terms of the structure of the state space, OR-graphs, AND/OR graphs and game trees are among the most well studied, among which, AND/OR graphs enable the modeling of alternative choices in conjunctive decompositions of the problem and its subproblems. Recently there has been a renewed research interest towards applying AND/OR structures in a several areas like planning with uncertainty [1], graphical models [2], web service composition [3], etc. In view of these recent research, the problem of generating ordered solutions becomes important for AND/OR graphs. Some of important applications include – (a) developing useful variants of the AO* algorithm, (b) graphical models, service composition [4], etc. The existing algorithms [5,4] work on explicit AND/OR graphs – thus the entire search space of the problem has to be made explicit before starting the algorithm. However, making the entire search space explicit is often not required for generating a reasonable number of ordered solutions. For several domains good heuristic functions are available and these estimates can be used to restrict the portion of the search space that is made explicit. We present a new algorithm, ASGAO*, which uses the heuristic estimates to dynamically make the search space explicit like AO* while computing the ordered solutions. Moreover, for problems like matrix chain multiplication, finding secondary structure of RNA, multi-peg tower of Hanoi, etc, whose search spaces have AND/OR structures, the optimal cost of subproblems can be computed efficiently. This exact

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 581–586, 2013.

value of the cost information can be used to further reduce the portions of AND/OR graph that is made explicit. Experimental results on domains like, including matrix chain multiplication, secondary structure of RNA, etc., reported in Section 4, demonstrate hundred times improvement in running time as well as significant improvement in space requirement over the existing algorithm on an average.

2 Background and Existing Approach

$G_{\alpha\beta} = \langle V, E\rangle$ denotes an AND/OR directed acyclic graph where V is the set of nodes, and E is the set of edges. The nodes of $G_{\alpha\beta}$ with no successors are called *terminal* nodes. The *non-terminal* nodes of $G_{\alpha\beta}$ are of two types – *i*) OR nodes and *ii*) AND nodes . The start (or root) node of $G_{\alpha\beta}$ is denoted by v_R. *OR edges* and *AND edges* are the edges that emanate from OR nodes and AND nodes respectively. We use the standard notion of *solution graph*, $S(v_q)$, rooted at any node $v_q \in V$ [6]. By a solution graph S of $G_{\alpha\beta}$ we mean a solution graph with root v_R. The notion of cost is as follows. In $G_{\alpha\beta}$, every edge $e_{qr} \in E$ from node v_q to node v_r has a finite non-negative cost $c_e(\langle v_q, v_r\rangle)$ or $c_e(e_{qr})$. Similarly every node v_q has a finite non-negative cost denoted by $c_v(v_q)$. We use the notion of additive costs for given *solution graph* S [6]. For every node v_q in S, $C(S, v_q)$ denotes the cost of the solution graph, $S(v_q)$. The cost of a solution S is $C(S, v_R)$ which is also denoted by $C(S)$. We denote the optimal solution below every node v_q as $opt(v_q)$. Therefore, the optimal solution of the entire AND/OR DAG $G_{\alpha\beta}$, denoted by S_{opt}, is $opt(v_R)$. The cost of the optimal solution rooted at every node v_q in $G_{\alpha\beta}$ is $C_{opt}(v_q)$. The cost of the optimal solution S_{opt} of $G_{\alpha\beta}$ is denoted by $C_{opt}(v_R)$ or, alternatively, by $C_{opt}(S_{opt})$. The *aggregated cost*, c_a, for an edge e_{ij} from node v_i to node v_j, is defined as : $c_a(e_{ij}) = c_e(e_{ij}) + C_{opt}(v_j)$

We briefly present the overview of the existing ASG algorithm for AND/OR trees. More details can be found in [4]. ASG algorithm works iteratively – starts by computing the optimal solution and subsequently generates the next best solution over iterations from the solutions that are already generated. In the context of AND/OR trees, e_q is used to denote the edge that points to the vertex v_q. ASG algorithm has been developed on the notion of – (a) marking of an OR edge, (b) swap option and swap operation and (b) the implicit representation of a solution. Consider a solution, S_{cur}, containing an OR edge $e_i = (v_q, v_i)$, where $S_{cur}(v_i)$ is actually $S_{opt}(v_i)$. Now, e_i is marked with the cost increment which will be incurred to construct the next best solution from S_{cur} by choosing another child of v_q, say v_j, and this difference is succinctly represented by a swap option $\sigma_{ij} = \langle e_i, e_j, \delta_{ij}\rangle$, where $\delta_{ij} = c_a(e_j) - c_a(e_i)$.

Swap operation basically denotes the application of a swap option to a solution. ASG use a swap option based compact representation, named *signature*, for storing the solutions. Intuitively, any alternative solution can be described as a set of swap operations performed on the optimal solution S_{opt}. The minimal sequence of swap options corresponding to a solution, S_m, is defined as the *signature*, $Sig(S_m)$, of that solution. The *swap list* corresponding to a solution S_m, $\mathcal{L}(S_m)$, is the list of swap options that are applicable to S_m. Let $Sig(S_m) = \{\sigma_1, \cdots, \sigma_m\}$ and $\forall i,\ 1 \leq i \leq m$, each swap option σ_i belongs to node v_{p_i}. The application of all other swap options that belong to the OR edges in the path from the root node to v_{p_i} is invalidated in the solution S_m.

Only the remaining swap options that are not invalidated in S_m can be applied to S_m for constructing the successor solutions of S_m. The set of successors of a solution S_m, denoted by $Succ(S_m)$, is the set $\{S'_m\}$ such that S'_m can be constructed from S_m by applying a swap option in $\mathcal{L}(S_m)$. These concepts are discussed in detail in [4].

ASG algorithm maintains a list, Open, which initially contains only the optimal solution S_{opt}. At any point of time Open contains a set of candidate solutions from which the next best solution in the non-decreasing order of cost is selected. At each iteration the minimum cost solution (S_{min}) in Open is removed from Open and added to another list, named, Closed which contains the set of ordered solutions generated so far. Then $Succ(S_{min})$ is constructed and any successor solution which is not currently present in Open as well as is not already added to Closed is inserted to Open.

3 Proposed Algorithm for Implicit AND/OR Graphs : ASGAO*

We use the underlying strategy of node expansion and cost revision of AO* repetitively in order to make the relevant portion of the AND/OR graph explicit. In the first step, ASGAO* computes the optimal solution using AO*. At this point of time portions of the AND/OR graph have been made partially explicit, i.e., some of the tip nodes of the AND/OR graph are terminal nodes, whereas the others are yet to expand.

Algorithm 1. Alternative solution generation for implicit AND/OR Graphs

input : $\mathsf{G}_{\alpha\beta}$ containing only the root node v_R and the state transformation rules
output : Ordered set of solutions
Run AO* and compute the optimal solution of $\mathsf{G}_{\alpha\beta}$;
Create two lists, Open and Closed, that are initially empty;
Compute the swap list, $\mathcal{L}(S_{opt})$, of S_{opt};
Construct $Succ(S_{opt})$ using $\mathcal{L}(S_{opt})$;
foreach $P_m \in Succ(S_{opt})$ **do** Add P_m to Open;
while Open *is not empty* **do**
    if P_{min} *is a solution* **then**
        P_{min} represents the next best solution S_{min};
        Add S_{min} to the Closed list and construct $Succ(S_{min})$;
        foreach $P' \in Succ(S_{min})$ **do**
            if $P' \notin$ Open **and** $P' \notin$ Closed **then** Add P_m to Open;
        end
    else
        Expand a tip node in P_{min} using the state transformation rules;
        Start bottom-up cost revision and adjust the swap options;
        Insert P_{min} back to Open;
end
Report the solutions in Closed;

We use $\mathsf{G}_{\alpha\beta}$ to denote a *partial explicit AND/OR graph* which initially consists of the only the root node v_R. The notion of cost for tip node is as follows. The cost of a tip node is returned by a heuristic function, denoted $h_{\alpha\beta}$. If a tip-node is not a terminal node, $h_{\alpha\beta}$ returns an estimate of the cost of solving the problem represented by that tip-node. On the contrary, when a tip-node is a terminal node, $h_{\alpha\beta}$ returns the cost of

the atomic subproblem represented by that terminal node. In this paper, we assume that the heuristic function is admissible and consistent. Unlike the ASG, ASGAO* works with potential solution graphs (psgs) [7] of $\mathsf{G}_{\alpha\beta}$. In a potential solution graph, every node with no child is a tip node. Swap options, swap lists, etc. can be computed on the partial explicit AND/OR graphs in likewise manner by considering the heuristic costs of the tip nodes. For partial explicit AND/OR graphs there are two types of swap options. Consider a swap option $\sigma_{pqr} = \langle e_{pq}, e_{pr}, \delta_{pqr} \rangle$ belonging to node v_p whose original edge is e_{pq} and the swapped edge is e_{pr}. If both of the optimal potential solution graphs rooted at v_q and v_r, do not have any tip node, then swap option σ_{pqr} is a *fixed* swap option, otherwise σ_{pqr} is a *changeable* swap option.

The first step of ASGAO* is to run AO* until it generates the optimal solution and after the computation of S_{opt}, the swap options are computed on the partial explicit graph, $\mathsf{G}_{\alpha\beta}$. AO* works over a number of iterations performing (*i*) computation of the current best psg, and (*ii*) expansion and cost update, till the optimal solution is found. Then the set of successor psgs of $Succ(S_{opt})$ are computed and added to Open. It is possible that some of the psgs can actually be a solution depending on whether all the tip nodes are terminal nodes. Next, ASGAO* performs the following over a sequence of iterations. First, the minimum cost psg, P_{min}, is removed from Open and there are the two possibilities. On the one hand, if P_{min} is a solution, it represents the next best solution S_{min}. ASGAO* adds S_{min} to the Closed list and constructs the successor set of psgs of S_{min}. Among the successor psgs, each psg which is not currently present in Open as well as is not already added to Closed is inserted to Open. On the other hand, if P_{min} is not an actual solution, P_{min} will contain some tip nodes which are not terminal nodes and one of them (chosen arbitrarily), say v_t, is selected and expanded using state transformation rules. If the cost of v_t is changed or its label is changed to solved due to the expansion, a bottom up cost revision step similar to AO* is triggered. In this cost revision step, apart from revising the cost, the swap options are also updated accordingly. After that, P_{min} is added back to Open. At any point of time, the Closed list contains the set of ordered solutions generated so far.

Correctness of ASGAO* : We present the key lemma and theorems to establish the correctness of ASGAO* under consistent and admissible heuristics. The details of the proofs are skipped for the sake of brevity.

Property 1. At any point of time during the execution of ASGAO*, the following holds.

1. For every solution S_{min} present in Closed, $Sig(S_m)$ contains no changeable swap option;
2. For every psg P_m present in Open, at most one changeable swap option can be present in $Sig(P_m)$;
3. While updating the swap options, the original edge of each changeable swap option which belongs to the signature of a psg, does not change. □

Lemma 1. *For an initial partial explicit AND/OR graph $\mathsf{G}_{\alpha\beta}$, for every solution of the corresponding fully explicit graph $\mathsf{G}^c_{\alpha\beta}$, ASGAO* generates that solution.* □

Lemma 2. *ASGSO* adds solutions to* Closed *in non-decreasing order of cost.* □

Table 1. Comparison of time required (in seconds) for AND/OR DAGs corresponding to the search spaces of *matrix-chain multiplication* with different number of matrices

$\#mat$	ASG				ASGAO*					
	$\#n$	10	20	30	$\#n$	10	$\#n$	20	$\#n$	30
20	1521	0.026	0.029	0.032	245	0.003	286	0.005	314	0.006
30	4931	0.179	0.186	0.194	506	0.006	517	0.009	539	0.012
40	11441	0.786	0.799	0.812	866	0.012	879	0.016	892	0.019
50	22051	2.590	2.612	2.637	1331	0.021	1344	0.026	1357	0.031
60	37761	6.847	6.878	6.940	1893	0.035	1900	0.041	1902	0.048
70	59571	16.003	16.104	16.212	5765	0.234	5945	0.275	5970	0.317
80	88481	33.968	34.187	34.408	4683	0.167	6368	0.322	6404	0.384
90	125491	66.220	66.601	66.989	5691	0.233	6915	0.380	7604	0.511
100	171601	120.326	121.042	121.605	6656	0.300	7880	0.464	8569	0.620

Table 2. Comparison of time required (in seconds) for AND/OR DAGs corresponding to the search spaces of *secondary structure of RNA*

Test Case	ASG				ASGAO*					
	$\#n$	10	20	30	$\#n$	10	$\#n$	20	$\#n$	30
TC1	63237	23.999	24.022	24.127	1333	0.039	1386	0.057	1476	0.089
TC2	84274	42.919	43.289	43.253	941	0.025	1399	0.058	1581	0.100
TC3	56649	19.631	19.732	19.801	1367	0.034	1682	0.054	2224	0.108
TC4	96484	53.132	53.280	53.552	1288	0.035	1408	0.054	1805	0.098
TC5	76787	33.669	33.811	34.028	1470	0.043	2296	0.085	3029	0.150
TC6	72397	31.667	31.884	32.053	1186	0.038	1927	0.074	2403	0.133
TC7	43471	10.708	10.764	10.829	789	0.019	1062	0.041	1253	0.069
TC8	49874	15.755	15.868	15.930	1055	0.030	1157	0.043	1296	0.063
TC9	45699	12.771	12.860	12.920	693	0.022	854	0.041	1145	0.079
TC10	108929	70.853	71.403	74.326	1093	0.032	1409	0.059	1549	0.093
TC11	73381	31.425	31.804	32.052	1106	0.032	1152	0.051	1404	0.081
TC12	50686	15.722	15.882	16.038	1146	0.032	1426	0.062	1512	0.087
TC13	90603	50.570	50.973	51.223	1254	0.035	1453	0.054	1773	0.095
TC14	56547	18.694	18.880	19.047	1382	0.037	1605	0.057	1980	0.095

Table 3. Comparison of time required (in seconds) for AND/OR Trees corresponding to the search spaces of *5-peg tower of Hanoi problem* with different number of disks

$\#disks$	ASG				ASGAO*					
	$\#n$	10	20	30	$\#n$	10	$\#n$	20	$\#n$	30
8	4941	0.019	0.020	0.021	101	0.004	148	0.007	157	0.009
9	17253	0.071	0.075	0.079	93	0.004	130	0.005	133	0.009
10	59049	0.227	0.246	0.261	99	0.005	102	0.006	197	0.010
11	199017	0.869	1.055	1.241	145	0.006	163	0.011	175	0.008
12	662661	3.101	3.767	4.436	215	0.009	288	0.016	302	0.018
13	2184813	10.742	12.764	14.949	267	0.011	410	0.018	436	0.021

Theorem 1. *For a given initial partial explicit AND/OR graph $\mathsf{G}_{\alpha\beta}$, every solution of the corresponding fully explicit graph is generated by ASGSO* only once and in the non-decreasing order of costs.* □

4 Experimental Results

To evaluate the performance of the ASGAO*, we have experimented on the following domains – (i) Matrix chain multiplication problem (ii) Secondary structure of RNA (iii) 5-peg tower of Hanoi problem and report the results in Table 1, Table 2 and Table 3 respectively. In these tables, $\#n$ denotes the number of nodes in the AND/OR graph and the time is reported for generating $10, 20,$ and 30 solutions. We also report the number of node present in the explicit graph used by ASG with the number of nodes expanded by ASGAO* for generating same number of ordered solutions. For all of these domains, we compute the exact optimal cost of solving the subproblem represented by any intermediate node using standard dynamic programming approach and use that value as heuristic. Due to the use of exact heuristic, ASGAO* does not invoke the cost revision step and significant improvement is achieved. The details of the test cases used for the secondary structure of RNA as well as the detailed description of each of the domains can be found in [4]. Table 1-3 show that when exact heuristic is used the size of the AND/OR DAG constructed by ASGAO* is significantly smaller than the size of the AND/OR DAG constructed by ASG. For matrix chain multiplication problem, we can observe around hundred times improvement in running time for bigger matrix chains and secondary structure of RNA. For generating 30 ordered solutions, the size of the explicit graphs are around five to twenty-five times larger compared to the size of the corresponding implicit graphs constructed by ASGAO* for matrix chain multiplication and secondary structure of RNA. For tower of Hanoi problem, the improvement is even more significant. Therefore, for domains like constraint satisfaction, planning etc., where good heuristics are available, ASGAO* can be used to generate ordered solutions efficiently, thereby opening up interesting research directions.

References

1. Russell, S., Norvig, P.: Planning. In: Artificial Intelligence: A Modern Approach, 2nd edn., pp. 375–461. Prentice-Hall, Englewood Cliffs (2003)
2. Dechter, R., Mateescu, R.: AND/OR search spaces for graphical models. Artif. Intell. 171(2-3), 73–106 (2007)
3. Lang, Q.A., Su, Y.: AND/OR graph and search algorithm for discovering composite web services. International Journal of Web Services Research 2(4), 46–64 (2005)
4. Ghosh, P., Sharma, A., Chakrabarti, P.P., Dasgupta, P.: Algorithms for generating ordered solutions for explicit AND/OR structures. J. Artif. Intell. Res (JAIR) 44, 275–333 (2012)
5. Elliott, P.: Extracting the k best solutions from a valued AND/OR acyclic graph. Master's thesis, Massachusetts Institute of Technology (2007)
6. Martelli, A., Montanari, U.: Additive AND/OR graphs. In: Proceedings of the 3rd International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers Inc., San Francisco (1973)
7. Dasgupta, P., Chakrabarti, P.P., DeSarkar, S.C.: Multiobjective heuristic search in AND/OR graphs. Journal of Algorithms 20(2), 282–311 (1996)

Intuitionistic Multi Fuzzy Soft Set and its Application in Decision Making

Sujit Das[1] and Samarjit Kar[2]

[1] Dept. of CSE, Dr. B.C. Roy Engineering College, Durgapur-713206, India
{sujit_cse,kar_s_k}@yahoo.com
[2] Dept. of Mathematics, National Institute of Technology Durgapur-713209, India

Abstract. Soft set theory initiated by Molodtsov in 1999 has been emerging as a generic mathematical tool for dealing with uncertainty. A noticeable progress is found concerning the practical use of soft set in decision making problems. This paper introduces the concept of intuitionistic multi fuzzy soft set (IMFSS) by combining the intuitionistic multi fuzzy set (IMFS) and soft set models. Then an algorithmic approach is presented by using induced fuzzy soft set and level soft set for dealing with decision making problem based on IMFSS. Finally the proposed algorithm has also been illustrated through a numerical example.

Keywords: Soft sets, intuitionistic fuzzy soft sets, intuitionistic multi fuzzy soft sets, decision making.

1 Introduction

In recent years a number of theories have been proposed to deal with uncertainty, imprecision and vagueness. Theory of probability, fuzzy set theory [1], intuitionistic fuzzy sets [2,3], vague sets [4], theory of interval mathematics [3,5], rough set theory [6] etc. are consistently being utilised as efficient tools for dealing with diverse types of uncertainties and imprecision embedded in a system. However, each of these theories has its inherent difficulties as pointed out by Molodtsov [7]. The reason for these difficulties is, possibly, the inadequacy of the parameterization tool of the theories. Molodtsov initiated a novel concept of soft set theory for modelling vagueness and uncertainty which is free from the difficulties affecting existing methods [7]. This theory has proven useful in many different fields such as decision making [8], data analysis [9], forecasting [10] and simulation [11].

Combining soft set models with other mathematical models has attracted the attention of many researchers. Maji et al. presented the concept of fuzzy soft set [12] which is based on a combination of the fuzzy set and soft set models. Yang et al. introduced the concept of the interval-valued fuzzy soft set [13] by combining the interval-valued fuzzy set and soft set. Feng et al. explored a relation between rough sets and soft sets, as also fuzzy sets and soft sets [14, 15]. By combining the trapezoidal fuzzy number and soft set, Xiao et al. proposed the concept of the trapezoidal fuzzy soft set [16]. Concept of multi fuzzy set [17] was proposed by Sebastian et al. which is a more general fuzzy set. Recently Shinoj el al. used the concept of intuitionistic

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 587–592, 2013.

multi fuzzy set for decision making purpose [18]. The purpose of this paper is to combine the intuitionistic multi fuzzy set and soft set, from which we can obtain a new soft set model named intuitionistic multi fuzzy soft set. To facilitate our discussion, we first review some preliminaries on soft set in section 2. In section 3, the concept of intuitionistic multi fuzzy soft set is presented. In section 4, intuitionistic multi fuzzy soft set is used to analyze a medical decision making problem using an algorithmic approach. Finally conclusions are pointed in section 5.

2 Preliminaries

Throughout this paper, U refers to an initial universe set, E is a set of parameters, P(U) is the power set of U and $A \subset E$.

Definition 1 [7]. A pair (F, A) is called a soft set over U, where F is a mapping given by $F : A \to P(U)$. In other words, a soft set over U is a mapping from parameters to P(U), and it is not only a set, but a parameterized family of subsets of U. For any parameter $e \in A, F(e)$ may be considered as the set of e-approximate elements of the soft set (F, A).

Definition 2 [12]. Let $\tilde{P}(U)$ denote the set of all fuzzy subsets of U, A pair $\left(\tilde{F}, A\right)$ is called a fuzzy soft set (FSS) over U, where $\tilde{F}$ is a mapping given by $\tilde{F} : A \to \tilde{P}(U)$.

Definition 3 [19, 20]. Let $I\hat{P}(U)$ denotes the set of all intuitionistic fuzzy sets of U. A pair $(\hat{F}, A)$ is called an intuitionistic fuzzy soft set (IFSS) over U, where $\hat{F}$ is a mapping given by $\hat{F} : A \to I\hat{P}(U)$. For any parameter $e \in A, \hat{F}(e)$ is an intuitionistic fuzzy subset of U and is called intuitionistic fuzzy value set of parameter e. Clearly $\hat{F}(e)$ can be written as an intuitionistic fuzzy set such that $\hat{F}(e) = \{< x, \mu_{\hat{F}(e)}(x), \nu_{\hat{F}(e)}(x), \pi_{\hat{F}(e)}(x) > | \, x \in U\}$. Here $\mu_{\hat{F}(e)}(x)$, $\nu_{\hat{F}(e)}(x)$ and $\pi_{\hat{F}(e)}(x)$ are membership and non membership functions and hesitation margin respectively and $\forall x \in U, \mu_{\hat{F}(e)}(x) + \nu_{\hat{F}(e)}(x) + \pi_{\hat{F}(e)}(x) = 1.$

Example: Assume that U is set of five diseases {Viral fever, Malaria, Typhoid, Stomach problem, Chest problem} given by $U = \{d_1, d_2, d_3, d_4, d_5\}$ and E is the set of symptoms {Temperature, Headache, Stomach pain, Cough, Chest pain} given by $E = \{s_1, s_2, s_3, s_4, s_5\}$. Let $A = \{s_1, s_2, s_3\}$. Then intuitionistic fuzzy soft set $\hat{(F, A)}$ can describe the possibilities of the diseases under intuitionistic fuzzy circumstances.

$\hat{F}(s_1) = \{d_1/(0.7,0.2,0.1), d_2/(0.8,0.2,0.0), d_3/(0.2,0.4,0.4), d_4/(0.3,0.2,0.5), d_5/(0.4,0.2,0.4)\}$,

$\hat{F}(s_2) = \{d_1/(0.7,0.2,0.1), d_2/(0.3,0.2,0.5), d_3/(0.9,0.0,0.1), d_4/(0.4,0.2,0.4), d_5/(0.1,0.2,0.7)\}$,

$\hat{F}(s_3) = \{d_1/(0.1,0.5,0.4), d_2/(0.8,0.2,0.0), d_3/(0.3,0.5,0.2), d_4/(0.7,0.1,0.2), d_5/(0.3,0.1,0.6)\}$.

Definition 4 [17]. An intuitionistic multi fuzzy set $\tilde{A}$ is defined by

$\tilde{A} = \{x/(\mu_A^1(x), \mu_A^2(x), ..., \mu_A^k(x)), (\nu_A^1(x), \nu_A^2(x), ..., \nu_A^k(x)), (\pi_A^1(x), \pi_A^2(x), ..., \pi_A^k(x)) : x \in U\}$,

such that $0 \le \mu_A^i(x) + \nu_A^i(x) + \pi_A^i(x) \le 1$ for every $x \in U$ and $i = 1, 2, ..., k$ where k is a positive integer and known as the dimension of $\tilde{A}$.

3 Intuitionistic Multi Fuzzy Soft Set

Definition 5. Let $IMFS(U)$ denotes the set of all intuitionistic multi fuzzy sets of U. A pair $(\mathcal{F}, A)$ is called a intuitionistic multi-fuzzy soft set (IMFSS) of dimension k over U, where $\mathcal{F}$ is a mapping given by $\mathcal{F} : A \to IMFS^k(U)$. An intuitionistic multi-fuzzy soft set is a mapping from parameters A to $IMFS^k(U)$. It is a parameterized family of intuitionistic multi fuzzy subsets of U. For $e \in A, \mathcal{F}(e)$ may be considered as the set of e-approximate elements of the intuitionistic multi fuzzy soft set $(\mathcal{F}, A)$.

In multi fuzzy set [17], membership grades by the experts are recorded at different time instants to evaluate the disease status. Intuitionistic multi fuzzy set provides the situation where each expert can prescribe his/her opinion in more accurate way using membership, non membership and hesitation margins. IMFSS $(\mathcal{F}, A)$ can describe the possibilities of the diseases using intuitionistic multi fuzzy sets. An IMFSS $(\mathcal{F}, A)$ can be easily represented by an Intuitionistic multi fuzzy soft matrix (IMFSM).

Definition 6. Using the definition of intuitionistic fuzzy soft matrix (IFSM) [21] we present the definition of intuitionistic multi fuzzy soft matrix to describe IMFSS.

Suppose $\mathcal{F} : E \to IMFS^k(U)$, $U = \{u_1, u_2, \cdots, u_m\}, E = \{e_1, e_2, \cdots, e_n\}$, the intuitionistic multi fuzzy soft matrix could be defined as $\mathcal{F} = (f_{ij})_{m \times n}$ where $f_{ij} = (\mu_{\mathcal{F}(e_j)}^l(x_i), \nu_{\mathcal{F}(e_j)}^l(x_i), \pi_{\mathcal{F}(e_j)}^l(x_i)), i = 1, 2, ..., m, j = 1, 2, ..., n, l = 1, 2, ..., k$,

$$[f_{ij}]_{m \times n} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{bmatrix}$$, where k is the dimension of IMFS

4 Application of IMFSS in Decision Making

4.1 Algorithmic Approach

Step 1. An intuitionistic multi fuzzy soft set $(\mathcal{F}, A)$ and relative weight $\omega(e_i)$ of parameter e_i is taken as input.

Step 2. Induced fuzzy soft set [22] $\Delta_{\mathcal{F}} = (\mathcal{F}, A)$ is computed.

Suppose $\omega(e) = (\omega_1, \omega_2, ..., \omega_n)^T$ where $\sum_{i=1}^{n} \omega_i = 1$ be the relative weight of parameter e, then induced fuzzy set can be defined as

$$[If_{ij}]_{m\times n} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{pmatrix} = \begin{pmatrix} \Sigma_{l=1}^{k} \{\omega(u_i).\mu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(v_i).\nu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(\pi_i).\pi^{l}_{\mathcal{F}(e_j)}(x_i) \\ \Sigma_{l=1}^{k} \{\omega(u_i).\mu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(v_i).\nu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(\pi_i).\pi^{l}_{\mathcal{F}(e_j)}(x_i) \\ \vdots \\ \Sigma_{l=1}^{k} \{\omega(u_i).\mu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(v_i).\nu^{l}_{\mathcal{F}(e_j)}(x_i) + \omega(\pi_i).\pi^{l}_{\mathcal{F}(e_j)}(x_i) \end{pmatrix}$$

Step 3. A threshold fuzzy set $\lambda : A \to [0,1]$ which is the mid level decision rule is chosen for decision making.

Step 4. The mid-level soft set $L(\Delta_{\mathcal{F}} : mid)$ is computed and presented in tabular form.

Step 5. The choice value c_i of u_i are calculated $\forall i$.

Step 6. The optimal decision is to select u_j if $c_j = \max(c_i) \forall i$.

Step 7. If j has more than one value, then any one of u_j might be chosen.

4.2 Experimental Analysis

Suppose a medical expert monitors the set of five symptoms {Temperature, Headache, Stomach pain, Cough, Chest pain} given by $E = \{s_1, s_2, s_3, s_4, s_5\}$ for a set of four patients $P = \{P_1, P_2, P_3, P_4\}$ to evaluate the most possible disease that a patient might suffer among a set of diseases {Viral fever, Malaria, Typhoid, Stomach problem, Chest problem} given by $U = \{d_1, d_2, d_3, d_4, d_5\}$. Observation of the expert is given in table 1. Let us assume that the medical system has imposed the following weights to the parameter set i.e. set of symptoms for the various possible diseases which are given in table 2. Table 3 shows the tabular representation of the induced fuzzy soft set for disease $d_1(\Delta_{\mathcal{F}d1} = (\mathcal{F}_{d1}, E)), \cdot d_2(\Delta_{\mathcal{F}d2} = (\mathcal{F}_{d2}, E))$ and $d_3(\Delta_{\mathcal{F}d3} = (\mathcal{F}_{d3}, E))$.

Mid-threshold of $\Delta_{\mathcal{F}_{d1}}$, $\Delta_{\mathcal{F}_{d2}}$ and $\Delta_{\mathcal{F}_{d3}}$ are fuzzy set given as follows.

$mid\,\Delta_{\mathcal{F}_{d1}} = \{(s_1, 1.07), (s_2, 1.05), (s_3, 0.93), (s_4, 1.07), (s_5, 1.11)\}$.

$mid\,\Delta_{\mathcal{F}_{d2}} = \{(s_1, 1.05), (s_2, 1.10), (s_3, 0.96), (s_4, 1.24), (s_5, 1.02)\}$.

Table 1. Expert's observation

Patients/ Symptoms	Temperature	Headache	Stomach pain	Cough	Chest pain
P_1	(0.3,0.7,0.5) (0.2,0.1,0.4) (0.5,0.2,0.1)	(0.4,0.3,0.4) (0.3,0.6,0.4) (0.3,0.1,0.2)	(0.1,0.2,0.0) (0.7,0.7,0.8) (0.2,0.1,0.2)	(0.5,0.6,0.7) (0.4,0.3,0.2) (0.1,0.1,0.1)	(0.4,0.3,0.4) (0.6,0.4,0.4) (0.0,0.3,0.2)
P_2	(0.4,0.3,0.5) (0.5,0.4,0.4) (0.1,0.3,0.1)	(0.7,0.6,0.8) (0.2,0.2,0.1) (0.1,0.2,0.1)	(0.6,0.5,0.4) (0.3,0.3,0.4) (0.1,0.2,0.2)	(0.3,0.6,0.2) (0.7,0.3,0.7) (0.0,0.1,0.1)	(0.4,0.7,0.5) (0.1,0.2,0.3) (0.5,0.1,0.2)
P_3	(0.1,0.2,0.1) (0.7,0.6,0.9) (0.2,0.2,0.0)	(0.3,0.2,0.1) (0.6,0.0,0.7) (0.1,0.8,0.2)	(0.8,0.7,0.8) (0.0,0.1,0.1) (0.2,0.2,0.1)	(0.3,0.2,0.2) (0.6,0.7,0.6) (0.1,0.1,0.2)	(0.4,0.3,0.2) (0.4,0.7,0.7) (0.2,0.0,0.1)
P_4	(0.5,0.4,0.5) (0.4,0.4,0.3) (0.1,0.2,0.2)	(0.4,0.3,0.4) (0.5,0.3,0.5) (0.1,0.4,0.1)	(0.2,0.1,0.0) (0.7,0.6,0.7) (0.1,0.3,0.3)	(0.5,0.6,0.3) (0.4,0.3,0.6) (0.1,0.1,0.1)	(0.4,0.5,0.4) (0.6,0.4,0.3) (0.0,0.1,0.3)

Table 2. Medical knowledgebase table

	Viral fever	Malaria	Typhoid	Stomach problem	Chest problem
Temperature	(0.5,0.3,0.2)	(0.6,0.2,0.2)	(0.4,0.1,0.5)	(0.1,0.7,0.2)	(0.1,0.8,0.1)
Headache	(0.6,0.1,0.3)	(0.3,0.6,0.1)	(0.3,0.1,0.6)	(0.2,0.4,0.4)	(0.2,0.5,0.3)
Stomach pain	(0.4,0.2,0.4)	(0.3,0.3,0.4)	(0.2,0.4,0.4)	(0.6,0.2,0.2)	(0.3,0.5,0.2)
Cough	(0.2,0.5,0.3)	(0.5,0.4,0.1)	(0.4,0.3,0.3)	(0.2,0.7,0.1)	(0.4,0.5,0.1)
Chest pain	(0.3,0.5,0.2)	(0.4,0.3,0.3)	(0.1,0.5,0.4)	(0.2,0.7,0.1)	(0.6,0.1,0.3)

Table 3. Induced fuzzy set

	Induced fuzzy set for disease d_1					Induced fuzzy set for disease d_2					Induced fuzzy set for disease d_3				
	s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5	s_1	s_2	s_3	s_4	s_5
P_1	1.12	0.97	0.76	0.90	1.13	1.20	1.17	0.95	1.29	1.01	1.07	0.82	1.14	1.08	1.01
P_2	1.09	1.43	1.00	1.13	0.94	1.08	0.97	0.95	1.25	1.06	0.86	0.92	0.90	1.01	1.78
P_3	0.94	0.82	1.16	1.21	1.23	0.76	1.07	0.95	1.15	0.99	0.58	0.97	0.74	0.97	1.11
P_4	1.13	0.97	0.80	1.02	1.12	1.16	1.17	0.97	1.25	1.03	0.92	0.82	1.14	1.04	0.94

Table 4. Mid level soft set

	Mid level soft set for d_1						Mid level soft set for d_2						Mid level soft set for d_3					
	s_1	s_2	s_3	s_4	s_5	Ch	s_1	s_2	s_3	s_4	s_5	Ch	s_1	s_2	s_3	s_4	s_5	Ch
P_1	1	0	0	0	1	2	1	1	0	1	0	3	1	0	1	1	1	**4**
P_2	1	1	1	1	0	**4**	1	0	0	1	1	3	1	1	0	0	0	2
P_3	0	0	1	1	1	3	0	0	0	0	0	0	0	1	0	0	1	2
P_4	1	0	0	0	1	2	1	1	1	1	1	**5**	1	0	1	1	0	3

$mid\Delta_{F_{d_3}} = \{(s_1, 0.86), (s_2, 0.88), (s_3, 0.90), (s_4, 1.03), (s_5, 0.96)\}$

Then we find the mid-level soft set $L(\Delta_{F_{di}} : mid)$ of $\Delta_{F_{di}}$ where $i = 1, 2, ..., 5$. Table 4 shows the mid level soft set and obtained choice values for the first three diseases. From table 4 it is obvious that patient P_2 has a possibility to suffer from viral fever, patient P_4 has a possibility to suffer from malaria and patient p_1 has a possibilty to suffer from typhoid. Calculating mid level soft set of $\Delta_{F_{d4}}$ and $\Delta_{F_{d5}}$ we find that patient P_3 has a possibility to suffer from stomach problem and chest problem might occur for both patients P_3 and P_4.

5 Conclusion

In this paper we propose the concepts of intuitionistic multi fuzzy soft sets, which is a combination of intuitionistic multi fuzzy sets and soft sets. This paper also introduces intuitionistic multi fuzzy soft matrix which will bring much convenience for future

researches on soft sets. Using induced fuzzy sets and level soft sets this paper uses an algorithmic approach to explore application of intuitionistic multi fuzzy soft sets in decision making problems. Finally we give an illustrative example to show the validity of the proposed approach. In future researchers might enhance this study to many other decision making problems that contain such kinds of uncertainties.

References

1. Zadeh, L.A.: Fuzzy sets. Inform. and Control 8, 338–353 (1965)
2. Atanassov, K.: Intuitionistic fuzzy sets. Fuzzy Sets and Systems 20, 87–96 (1986)
3. Atanassov, K.: Operators over interval valued intuitionistic fuzzy sets. Fuzzy Sets and Systems 64, 159–174 (1994)
4. Gau, W.L., Buehrer, D.J.: Vague sets. IEEE Trans. System Man Cybernet. 23(2), 610–614 (1993)
5. Gorzalzany, M.B.: A method of inference in approximate reasoning based on interval-valued fuzzy sets. Fuzzy Sets and Systems 21, 1–17 (1987)
6. Pawlak, Z.: Rough sets. Internat. J. Inform. Comput. Sci. 11, 341–356 (1982)
7. Molodtsov, D.: Soft set theory – first results. Comput. Math. Appl. 37(4-5), 19–31 (1999)
8. Maji, P.K., Roy, A.R., Biswas, R.: An application of soft sets in a decision making problem. Comput. Math. Appl. 44(8-9), 1077–1083 (2002)
9. Zou, Y., Xiao, Z.: Data analysis approaches of soft sets under incomplete information. Knowledge-Based Syst. 21(8), 941–945 (2008)
10. Xiao, Z., Gong, K., Zou, Y.: A combined forecasting approach based on fuzzy soft sets. J. Comput. Appl. Math. 228(1), 326–333 (2009)
11. Kalayathankal, S.J., Singh, G.S.: A fuzzy soft flood alarm model. Math. Comput. Simul. 80(5), 887–893 (2010)
12. Maji, P.K., Biswas, R., Roy, A.R.: Fuzzy soft sets. J. Fuzzy Math. 9(3), 589–602 (2001)
13. Yang, X.B., Lin, T.Y., Yang, J.Y., Li, Y., Yu, D.Y.: Combination of interval-valued fuzzy set and soft set. Comput. Math. Appl. 58, 521–527 (2009)
14. Feng, F., Liu, X.Y., Fotea, V.L., Jun, Y.B.: Soft sets and soft rough sets. Inform. Sci. 181, 1125–1137 (2011)
15. Feng, F.C., Li, B., Davvaz, A.M.I.: Soft sets combined with fuzzy sets and rough sets: a tentative approach. Soft Comput. 14, 899–911 (2010)
16. Xiao, Z., Xia, S., Gong, K., Li, D.: The trapezoidal fuzzy soft set and its application in MCDM. Appl. Math. Model (2012), http://dx.doi.org/10.1016/j.apm.2012.01.036
17. Sebastian, S., Ramakrishnan, T.V.: Multi-fuzzy sets: an extension of fuzzy sets. Fuzzy Inform. Eng. 1, 35–43 (2011)
18. Shinoj, T.K., John, S.J.: Intuitionistic Fuzzy Multisets And Its Application in Medical Diagnosis. World Academy of Science, Engineering and Technology 61, 1178–1181 (2012)
19. Maji, P.K., Biswas, R., Roy, A.R.: Intuitionistic fuzzy soft sets. J. Fuzzy Math. 9(3), 677–692 (2001)
20. Maji, P.K., Roy, A.R., Biswas, R.: On intuitionistic fuzzy soft sets. J. Fuzzy Math. 12(3), 669–683 (2004)
21. Mao, J., Yao, D., Wang, C.: Group decision making methods based on intuitionistic fuzzy soft matrices. Applied Mathematical Modelling 37, 6425–6436 (2013)
22. Yang, Y., Tan, X., Meng, C.: The multi fuzzy soft set and its application in decision making. Applied Mathematical Modelling 37, 4915–4923 (2013)

Incorporation of Particle Swarm Optimization in Adaptive Boosting

Gaurav Mishra, Rohit Kumar, and Santanu Chaudhury

Electrical Engineering Department, Indian Institute of Technology, Delhi, India

Abstract. This paper proposes an optimized learning method for large feature-sets using AdaBoost to produce hardware-efficient boosted decision stumps. The paper also proposes a method for training decision stumps to construct the ensemble. AdaBoost sequentially searches for the best weak classifier in the pool and adds it to the ensemble, using weighted training samples. In the proposed method, Particle Swarm Optimization quickens the selection of decision stumps. It is shown experimentally that the optimized method is more than 60% faster than the exhaustive search method.

1 Introduction

An ensemble of weak classifiers is often used to produce a strong classifier[9]. AdaBoost[6, 10] provides a set of methods for combining weak classifiers to construct ensembles. Adaboost facilitates the sequential combination of weak classifiers to produce a strong ensemble, but does not address the optimal selection of these classifiers. In this paper, we propose a method that incorporates Particle Swarm Optimization[7] into AdaBoost, for the optimal selection of weak classifiers. A method to efficiently train weak decision stumps is also proposed, as these are easily implementable on hardware[4].

Zhu et al. [2] and Zhang et al. [3] used AdaBoost to select features from histogram of oriented gradients[8], especially for human detection. Viola et al.[1] used AdaBoost to select an ensemble of decision stumps from a pool of weak stumps. In such a framework, training a large pool of stumps and their optimal selection is an important issue, which is addressed in this paper.

Evans et al. [5] used Particle Swarm Optimization (PSO) [7] to find bounds on classes with respect to all features. The image features created and used in the Viola-Jones framework are such that a gradient-descent approach can be used to find local minimas in error rates produced by using single features for training. Hence the AdaBoost procedure is suitable for and can be enhanced by PSO.

However, PSO has not been used for the design of cascaded boosted classifiers. The rest of the paper is organized in the following manner. The proposed method is described in Section 2. Experimental validation is provided in Section 3, and Section 4 concludes.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 593–598, 2013.

2 Optimized Boosting of Decision Stumps

2.1 Classifier Selection Using Adaboost

An ensemble of weak classifiers can be constructed using AdaBoost, by sequentially selecting one classifier from the pool of weak classifiers and redistributing the weights of samples. In each step, the pool of classifiers is tested on the weighted training set, rated according to their performance, and the best classifier is selected. So, for large pools, exhaustive search becomes computationally expensive. AdaBoost does not address the issue of optimal selection of classifiers.

In this paper, we shall propose a method to accelerate the selection procedure of AdaBoost, using Particle Swarm Optimization. Particle Swarm optimization initializes a set of particles (classifiers from the pool, in this case) and based on their accuracy, converges towards the best weak classifier in the pool. We also propose an optimized method to create the pool of decision stumps.

2.2 Proposed Method

Algorithm 1. AdaBoost

1. **Input** D-dimensional training samples$S = (\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), ...(\boldsymbol{x_m}, y_m)$ (where $\boldsymbol{x_i} = (x_i^1, x_i^2, ..., x_i^d)^T$ is the i^{th} sample, y_i is the class of i^{th} sample), number of iterations P, number of particles N, number of PSO iterations T
2. Train D decision stumps $\boldsymbol{h} = (h_1, h_2, ..., h_D)$, one with respect to each feature $h_d : \boldsymbol{x} \mapsto \{-1, +1\}$
3. Initialize weights $w_i^0 = 1/M$ for all $i = 1...M$
4. **Do** for $p = 1...P$:
   (a) Select optimal classifier h^p using PSO, following Algorithm 2
   (b) Calculate weighted error of h^p: $e^p = \Sigma_{i=1}^{M} w_i^p I(y_i \neq h^p(x_i))$, and coefficient $\alpha^p = \frac{1}{2}\log(\frac{1-e^p}{e^p})$
   (c) Update the weights: $w_i^p = \frac{w_i^{p-1}}{Z_p}\exp\{-\alpha_p y_i h^p(x_i)\}$, where $Z_p = 2\sqrt{e^p(1-e^p)}$
   (d) Stop if $e^p = 0$ or $e^p > 1/2$ and set $T = t - 1$
5. **Output** $y(x) = \text{sgn}(\Sigma_{p=1}^{P}\alpha^p h^p(\boldsymbol{x}))$

Algorithm 2. Optimal Selection Using PSO

1. **Input** samples $S = \{(\boldsymbol{x_1}, y_1, w_1^p), (\boldsymbol{x_2}, y_2, w_2^p), .(\boldsymbol{x_M}, y_M, w_M^p)\}$, number of particles N, number of PSO iterations T, pool of decision stumps $\boldsymbol{h} = (h_1, h_2, ..., h_D)$
2. Randomly select N features $k^0 = (k_1^0, k_2^0, ..., k_N^0)$ and corresponding decision stumps $H^0 = (h_{k_1^0}, h_{k_2^0}, ..., h_{k_N^0})$
3. Initialize velocities $V_n^0 = 0$ corresponding to each $h_{k_n^0}$ for $n = 1...N$
4. Calculate weighted errors $e_n = \Sigma_{i=1}^{M} w_i^p I(y_i \neq h_{k_n^0})$
5. Initialize global best particle $k^* = k_n^0 : \{n = \text{argmin}_n e_n\}$, and corresponding global best decision stump $h^* = h_{k^*}$

6. Initialize local best particles $lk_n^0 = k_n^0$, local best decision stumps $l_n^0 = h_{lk_n^0}$, and local best errors $le_n^0 = e_n^0$,for $n = 1...N$
7. **Do** for $t = 1...T$:
   (a) Update velocities:

$$V_n^{t+1} = r_1 V_n^t + r_2(lk_n^t - k_n^t) + r_3(k^* - k_n^t) \quad (1)$$

   (b) Update features and take corresponding decision stumps:

$$k_n^{t+1} = k_n^t + V_n^t \text{ for } n = 1...N \quad (2)$$

   If $k_n^t > D$ then $k_n^t = D$, If $k_n^t < 1$ then $k_n^t = 1$
   (c) Calculate weighted errors $e_n = \Sigma_{i=1}^M w_i^p I(y_i \neq h_{k_n^t}(\boldsymbol{x_i}))$
   (d) Update local best stumps and errors:

$$lk_n^t = \begin{cases} lk_n^{t-1} & \text{if } e_n^t \geq le_n^t \\ k_n^t & \text{otherwise} \end{cases}, l_n^t = h_{lk_n^t} \text{ and } le_n^t = \min(le_n^t, e_n^t) \quad (3)$$

   (e) Update global best stump $h^* = \{l_n^t : n = \text{argmin}_n le_n^t\}$
8. **Output** h^*

In the above algorithm, r_1, r_2 and r_3 are the constants used to control the movement of the particles. Empirical studies have recommended the values $r_1 = 0.729844$, $r_2 = r_3 = 1.4961798$ for the optimal convergence of PSO[11]. Van den Berg[12] has shown that PSO is locally convergent. Hence, if enough particles are initailized over the search space, the probability of reaching the global minima is high. Empirical analysis has shown PSO to be reliable optimization technique[7].

2.3 Training Decision Stumps

To train decision stumps described in the proposed method, a similar approach using PSO can be used. For example, consider a 256×256 2-D sample space. Suppose we select only integer stumps, then a threshold has to be selected from 256 possibilities w.r.t. one feature. Instead of an exhaustive search, an optimized search using PSO can be executed.

3 Experimental Validation

3.1 Datasets Used

The proposed method was tested on 3 databases, the INRIA pedestrian database[1], a database of faces[2], and the Gissette database[3] (see Table 1). The INRIA dataset

[1] Available at `www.pascal.inrialpes.fr/data/human/`
[2] Available at `cswww.essex.ac.uk/mv/allfaces/index.html`
[3] Available in the UCI repository[13], `www.archive.ics.uci.edu/ml/datasets/Gisette`

provides images of pedestrians. The Face database contains images of faces to use for training. HOG features were extracted from these images to train the decision stumps. For the other datasets, extracted features were available and were used for training purposes. Decision stumps were trained in the following manner: for each feature, the mean value of both classes was taken, and the mean of these two values was taken as the threshold.

Table 1. Description of Datasets

Dataset	Number of Samples	Number of features
INRIA	3634	1610
Faces	6078	4140
Gisette	6000	5000

The time required for training and the accuracy of resulting ensembles using exhaustive search and optimized search were compared using these feature sets.

3.2 Results

The overall error as more features are added to the ensemble is plotted w.r.t. training time in Fig. 1,2, and 3. The overall classification error decreases exponentially as more weak classifiers are added[6]. Clearly, the optimized search takes less time construct an ensemble than exhaustive search. The training time taken to achieve 10% error rate, 5%, 2% and zero classification error are tabulated in Table 2. The improvement in training times using PSO search are tabulated in Table 3. In most cases, an improvement of more than 60% is observed.

The training method for decision stumps using PSO was tested using a sample 2D dataset shown in Fig. 4. An ensemble of 50 decision stumps is shown in Fig. 5 and of 200 decision stumps is shown in Fig. 6. Using more stumps gives a closer approximation of the linear boundary between the classes.

Table 2. Training Time(sec)

Dataset	10% error		5% error		2% error		0 error	
	Exhaustive	PSO	Exhaustive	PSO	Exhaustive	PSO	Exhaustive	PSO
INRIA	4	1	13	6	49	21	274	106
Face	19.7	3.8	45.6	15.2	175.4	60.9	1501.8	425.7
Gisette	21	8	108	41	506	198	1970	777
Car	5.6	0.8	20.8	8.4	-	-	-	-

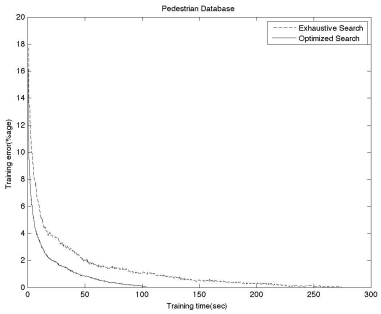


Fig. 1. Pedestrian: Training Error(%age) vs. Training Time(sec)

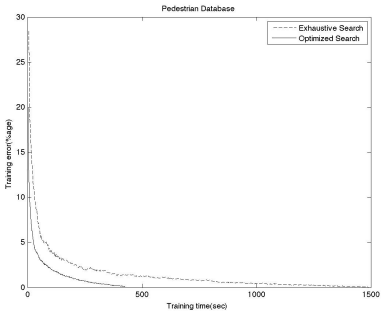


Fig. 2. Faces: Training Error(%age) vs. Training Time(sec)

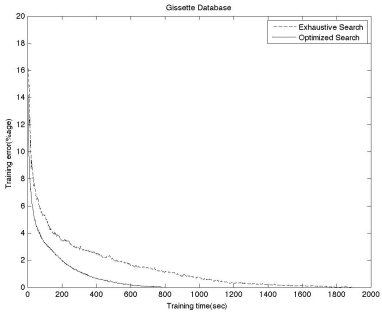


Fig. 3. Gissette: Training Error(%age) vs. Training Time(sec)

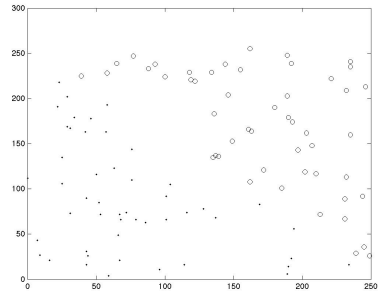


Fig. 4. Sample 2D Dataset

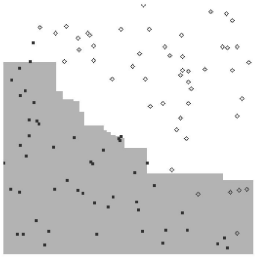


Fig. 5. Training Decision Stumps: 50 stumps

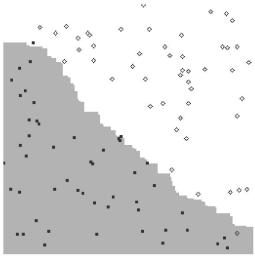


Fig. 6. Training Decision Stumps: 200 stumps

Table 3. Improvement in Training Time(%age)

Dataset	10%error	5%error	2%error	0 error
INRIA	75	53.84	57.14	61.31
Face	80.71	66.67	65.27	71.65
Gisette	61.9	62.03	60.86	60.55
Car	85.71	59.61	-	-

4 Conclusion

In this paper we presented a modified approach to Adaptive Boosting for object detection applications. Particle Swarm Optimization was used to improve the time taken to construct ensembles. It was demonstrated that this reduction was more than 60%. This method can be used to produce ensembles in a faster and more efficient manner.

References

[1] Viola, P., Jones, M.: Robust Real-time Object Detection. International Journal of Computer Vision (2001)
[2] Zhu, Q., Avidan, S., Yeh, M., Cheng, K.: Fast human detection using a cascade of histograms of oriented gradients. In: CVPR 2006, pp. 1491–1498 (2006)
[3] Jia, H., Zhang, Y.: Fast Human Detection by Boosting Histograms of Oriented Gradients. In: Fourth International Conference on Image and Graphics, pp. 683–688 (2007)
[4] Miteran, J., Matas, J., Bourennane, E., Paindavoine, M., Dubois, J.: Automatic hardware implementation tool for a discrete adaboost-based decision algorithm. EURASIP Journal on Applied Signal Processing, 1035–1046 (2005)
[5] Evans, H., Zhang, M.: Particle swarm optimisation for object classification. In: 23rd International Conference on Image and Vision Computing, New Zealand, pp. 1–6 (2008)
[6] Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of on-Line Learning and An Application to Boosting (1995)
[7] Kennedy, J., Eberhart, R.: Particle swarm optimization Proceedings. In: IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948 (1995)
[8] Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: CVPR, pp. 886–893 (2005)
[9] Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley-Interscience (2004)
[10] Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. In: Proceedings of the 11th Annual Conference on Computational Learning Theory, pp. 80–91. ACM (1998)
[11] Eberhart, R.C., Shi, Y.: Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. In: Proceedings of the IEEE Congress on Evolutionary Computation, San Diego, USA (2000)
[12] Bergh, F., Engelbrecht, A.P.: A Convergence Proof for the Particle Swarm Optimiser (2010)
[13] Guyon, I., Gunn, S.R., Ben-Hur, A., Dror, G.: Result analysis of the NIPS 2003 feature selection challenge (2004)

Exploring a Quantum Hebbian Model of Feature Map Formation

Priti Gupta and C.M. Markan

Department of Physics and Computer Science, Dayalbagh Educational Institute, Agra, India
{gupta.priti.84,cm.markan}@gmail.com

Abstract. The brain self-organizes into feature maps or neural assemblies on receiving inputs. Similar self-organization is possible in artificial systems only if the principles that the brain employs are exploited. Existing models explaining feature map formation cover only some aspects of local feature map formation. It is unlikely that the brain employs different mechanisms to form local and global feature maps and hence there is a need to explore a single mechanism that could account for neural interactions at all levels. If we take the brain to coexist as a quantum and a classical system, certain insights can be obtained about neural development. In this paper we explore a quantum hebbian model of interaction between the quantum and classical processes in the brain, which in synergy with mental force of directed attention, seems to have the potential to explain the formation of both local and global feature maps. Introducing this duality also helps us address higher level issues like mind wandering, zombie modes, volition etc.

1 Introduction

The brain's cortical structure adapts to environmental inputs and hardwires itself during a critical learning period [15]. If we can have machines that adapt to the environment in a similar way, we move a step closer to having truly intelligent machines. The brain organizes itself into feature maps in response to the inputs it receives. If we can replicate such feature maps in hardware, it could have wide applications in robotics, artificial sensory systems and neural prosthetics. Recently, it has been shown that analogue hardware can be fine-tuned to work in the environment for which it is designed by introducing field programmability [7]. While technological advances have paved the way for us to build neuromorphic hardware that can emulate feature map formation [3], what is important now is to understand in depth how global feature maps are formed in the brain. For simplification feature maps can be broadly classified into two types: local and global feature maps. While local feature maps represent the connectivity patterns within a localized cortical area e.g. visual, auditory etc., global feature maps represent the connectivity patterns across spatially separated cortical areas. Existing models [2] assume prior interconnectivity in the form of Mexican Hat connections. This kind of precise connectivity is not found in a developing

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 599–604, 2013.

brain. Other models, based on non-axonal means of communication e.g. Diffusive Hebbian [9] etc, overcome this limitation, but they are confined to intra-modality or local interaction. It is unlikely that the brain employs different mechanisms for forming local and global connections and therefore this problem needs re-examination. Neural connections are formed by the famous Hebb's Rule 'Cells that fire together wire together' and Stent's modification 'Cells that do not fire together compete for elimination' i.e. if the activity of two neurons is synchronized, synaptic connection joining them strengthens whereas if the activity is not synchronized this strength decreases and finally the connection is eliminated. To understand how inter-cortical or global neural assemblies are formed, we need to decipher how in the absence of any direct physical connections, spatially separated neurons synchronize (with zero-lag)? This is a paradoxical situation that we term as the Neural Development Paradox (NDP). In order to understand how global feature maps are formed we need to break the NDP.

2 Need for a Quantum Mechanical Approach

The motivation to use quantum mechanics to unravel some of the mysteries of the brain stems from many reasons. Biological systems are made up of molecules and atoms and the most advanced theory that explains the interaction between them is quantum theory. Recently, quantum effects have been shown to operate at high temperatures in biological systems e.g. in birds vision, photosynthesis, olfaction [8] and many experiments are being performed to find out if they exist in the brain [5]. The brain processes various sensory inputs e.g. visual, auditory etc. at different times and locations, even then we perceive them as simultaneous events. Since synaptic transmission and axonal transfer of nerve impulses are too slow to organize coordinated activity in large areas of the brain, other non-axonal means need to be explored. In a recent theory, Hameroff et.al. [4, 5] suggest that Microtubules inside cells support long-range quantum coherence, enabling quantum information processing to take place at the sub-cellular level. Cells interconnected by gap junctions form networks which fire synchronously, behaving like one giant neuron [6], and possibly accounting for synchronized neural activity such as coherent 40 Hz waves, and could be a possible solution to the NDP.

3 Formation of Neural Assemblies and Quantum Zeno Effect

The formation of neural assemblies is based on three events. First, for a particular stimulus, a single neural assembly is activated from all possible combinations of neural assemblies in the brain. Second, there is sustained synchronous activity of the neurons in that assembly, and third, there is change in the synaptic weights connecting its neurons, so that they become hardwired. All these events can be explained in quantum mechanical parlance by considering Quantum Entanglement, Quantum Zeno Effect (QZE) [11] and Quantum Decoherence. QZE

states that we can maintain the state of a system as long as we perform measurement on it. According to Penrose-Hameroffs quantum approach, the subneuronal microtubules form a quantum computer whose collapsed states are conscious experiences manifesting in the form of an activated neural assembly. Quantum Entanglement among the sub neuronal microtubules of neurons [4] are responsible for the selection of a group of neurons forming a neural assembly. The neurons whose microtubules are entangled get connected by gap junctions to form one large hyper-neuron. When the quantum wave function collapses or de-coherence occurs, after a critical threshold has been reached [1], the neurons in the neural assembly fire in synchrony. Quantum Zeno Effect sustains this synchronous activity by repeated collapses to the same state of the system so that hebbian learning can strengthen the synaptic connections. The question is what causes the QZE? The role of attention in the learning process has always been acknowledged. The relation between attention and QZE can be seen in [13, 14] and it seems reasonable to establish that attention causes the Quantum Zeno effect to maintain the brain in a particular state and facilitates neuro-plastic changes.

4 Quantum Neural Computation

As a preliminary effort towards understanding how quantum computation in the brain works, we build a basic unit called a Quantum Neural Computer (QNC). Previous efforts to combine the fields of Quantum Computing and Artificial Neural Networks [12] have been made to enhance the computational power of ANNs, but none has been used with a view to understand brain function. In our model, the function of the QNC is analogous to the microtubule cytoskeleton quantum computer. The QNC takes N inputs that are either $|0>$ or $|1>$ and transforms them to an output state $u = \alpha\,|0> + \beta\,|1>$ where α and β are derived from the weights and α^2 and β^2 are the probabilities that the output u will be a $|0>$ or $a\ |1>$. The unitary transformation is implemented by a weight matrix described in [10] given as:

$$W_i = \sum_{x=0}^{2^N-1} c_x^i(x) \qquad where\ c_x^i \,\epsilon\, \{0,1\}\ and\ x\,\epsilon\, I \tag{1}$$

For the sake of normalization, only one c_x^i is 1 for a particular input, the rest are 0. The α and β are computed as below.

$$\alpha = \sum_i \delta_{a_{i,0}}\ W_i\,, \quad \beta = \sum_i \delta_{a_{i,1}}\ W_i \qquad and \quad \alpha^2 + \beta^2 = 1 \tag{2}$$

There are $2^3 \times 2^3 \times 2^3$ possible weight matrices. The strength of quantum computation lies in these many possibilities of transformations and this is what we harness in our model. The QNC computes all possible output states, transforming the inputs over all possible weight matrices simultaneously, by means of superposition. On collapse, the system reduces to a single output state that is chosen randomly from the set of all possibilities.

5 Quantum Hebbian Learning

Every Quantum collapse leads to increment in knowledge [14]. In the human brain, the modification in synaptic weights accounts for accumulation of knowledge. In order to understand how classical synaptic weights can be affected by quantum collapses we need to model their interaction. Let C be a certain classical weight, δ be the classical learning rate (representing the classical accumulation of knowledge for every quantum collapse), and Q to be the impact of the current state of QNC. δ is a very small value (say .001). By repeated collapses to the same state by means of attention and QZE, the classical learning becomes sufficiently large, causing a permanent change in the synaptic weights. The equation below describes the above interaction.

$$Next\,Collapsed\,State \leftarrow \ \alpha Q + (1 - |\alpha Q|)^* C \tag{3}$$

Where, $C = C + \delta Q$, α*:* attention parameter, δ: classical learning rate, Q: current state of the QNC, αQ: qualia, C: classical weight. Here the first part of the equation αQ determines the quantum influence and the term $(1 - |\alpha Q|)^* C$ determines the classical influence on the next collapsed state. Depending on the values of the various parameters that are involved in this equation, many interesting situations arise(see table 1).

Mind Wandering: The absence of attention ($\alpha = 0$) and when the brain is not developed (i.e. C is very small), leads to a state where all collapses of the QNC are random (determined by the quantum influence). This is analogous to the state of mind wandering.

Table 1. (Here, $\beta = (\,1 - (\,|\,\alpha Q\,|\,))$

	Non Developed Brain (C is low)	*Developed Brain (C is high)*
No Attention ($\alpha = 0$)	Mind Wandering	Zombie
Full Attention ($\alpha = 1$)	Quantum Zeno (when αQ > βC)	Volitional Choice (when αQ > βC)
	Mind Wandering (when αQ < βC)	Zombie (when αQ < βC)

Zombie: In absence of attention ($\alpha = 0$), when the brain is developed (C is large), the classical accumulated knowledge determines the state to which the subsequent collapse happens e.g. when youre driving a car in an auto-pilot mode without paying attention to details.

Quantum Zeno: In the presence of attention ($\alpha = 1$) and when the classical weights have not developed i.e. $\alpha Q > \beta C$, by focused attention the subsequent collapses can be made to happen to the same state. The quantum part plays the main role. In this state accumulation of knowledge takes place, $C = C + \delta Q$. When C become sufficiently large, we say hebbian learning has strengthened the synaptic connections of the assembly.

Volitional Choice (Free Will): This is a situation that arises when even though the classical weights are well developed, by keeping attention high, the collapse of the QNC can be made to be determined by the quantum part rather than the classical part. This bears analogy with the problem of volitional choice which has not been explained neurobiologically. It seems that by having a dual system (quantum + classical) in which one system has a supervisory role, some insight into how free will could be working can be obtained. A MATLAB simulation model of the Quantum Hebbian paradigm was developed for a 3 input QNC. Figure 1 shows the simulation results.

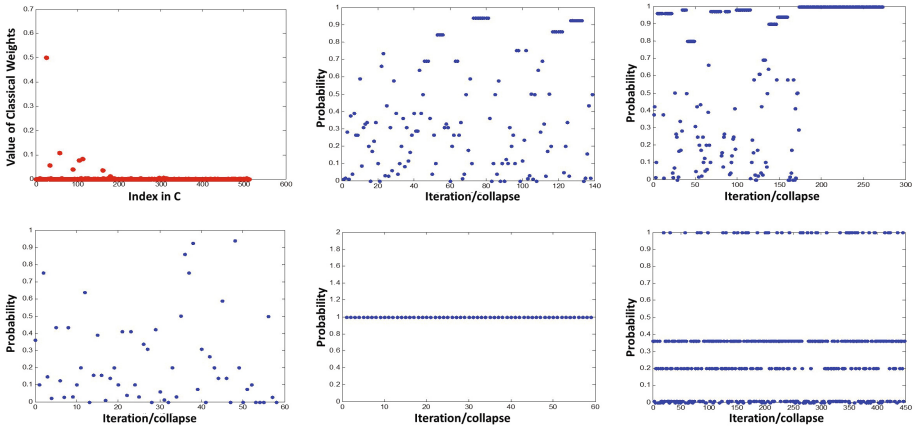


Fig. 1. [Row 1, a-c, row 2, d-f] a). Accumulation of classical weights. Out of many possibilities of neural assemblies, weights of only one neural assembly have been strengthened (the x-axis shows the index of the classical weights and y-axis shows the value of classical weight). b). QZE, repeated collapses to the same state can be observed from the continuous dots. c). Mind wandering state($\alpha = 1$). Collapses are mostly random, QZE too small to change classical weights permanently. d). Mind wandering state($\alpha = 0$), the collapses are completely random. e). Shows the zombie effect,($\alpha = 0$), the collapses are determined by the classical weights that have developed the most. f). Free will state. Here despite of the presence of a developed classical weight (dots corresponding to probability = 1), majority of the collapses are happening to a state determined by the quantum part (dots corresponding to probability .35) (for figures b-f, x-axis represents the number of iteration and y-axis shows the probability)

6 Discussion

A framework for quantum and classical interaction in the brain has been proposed. The Quantum Hebbian Equation not only indicates how quantum collapses could lead to classical learning but also determines to which state subsequent collapses will happen. With the attention parameter regulating the QNC, insight into the behaviour of the brain in the presence and absence of attention can be obtained. The different states of the brain like learning in presence of attention, zombie and mind wandering and volitional choice have been explained on the basis of this quantum and classical interaction. While this work is mostly exploratory in the present state, with research in quantum computing and VLSI design on the fore, it is just a matter of time that we would be able to create machines based on this kind of dual interaction of quantum and classical information.

References

1. Diosi, L.A.J.O.S.: Models for universal reduction of macroscopic quantum fluctuations. Physical Review A 40(3), 1165 (1989)
2. Erwin, E., Obermayer, K., Schulten, K.: Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. Neural Computation 7(3), 425–468 (1995)
3. Gupta, P., Bansal, M., Markan, C.M.: Neuromorphic Adaptable Ocular dominance maps. In: Ghosh, A., De, R.K., Pal, S.K. (eds.) PReMI 2007. LNCS, vol. 4815, pp. 49–56. Springer, Heidelberg (2007)
4. Hameroff, S.R.: Quantum coherence in MTs: a neural basis for emergent consciousness? Journal of Consciousness Studies 1, 98–118 (1994)
5. Hameroff, S., Tuszynski, J.: Quantum states in proteins and protein assemblies: The essence of life? In: Proceedings of SPIE Conference on Fluctuationa and Noise, Canary Islands (2004)
6. Kandel, E.R., Siegelbaum, S.A.K., Schwartz, J.H.: Synaptic transmission. In: Kandel, E.R., Schwartz, J.H. (eds.) Principles of Neural Science, 3rd edn., pp. 121–134. Elsevier, New York (1991)
7. Kapur, G., Mittal, S., Markan, C.M., Pyara, V.P.: A Unique Design methodology to generate reconfigurable Analog ICs with simplified Design Cycle. In: Workshop on Unique Chips and Systems UCAS-7, p. 28 (2012)
8. Lambert, N., Chen, Y.N., Cheng, Y.C., Li, C.M., Chen, Y.G., Nori, F.: Quantum Biology. Nature Physics 9, 10–18 (2012)
9. Markan, C.M., Bhaumik, B.: A Diffusive-Hebbian model for cortical orientation maps formation. In: Proc. of IJCNN 1999, Washington DC (1999)
10. Miranker, W.L.: Quantum Neurons. Tech report, Univ. of Yale, Yale/DCS/tr1234 (2002)
11. Misra, B., Sudarshan, E.C.G.: The Zeno's paradox in quantum theory. Journal of Mathematical Physics 18, 756 (1977)
12. Rudolph, T.G., Jensen, H.J.: A Heuristic Review of Quantum Neural Networks (2011)
13. Schwartz, J.M., Begley, S.: The mind and the brain. HarperCollins e-books (2009)
14. Stapp, H.P.: Mind, matter and quantum mechanics. Springer, Berlin (1993)
15. Sur, M., Leamy, C.A.: Development and Plasticity of Cortical Areas and Networks. Nature 2, 251–262 (2001)

Real Parameter Optimization Using Levy Distributed Differential Evolution

Nanda Dulal Jana[1], Aditya Narayn Hati[1], Rajkumar Darbar[2], and Jaya Sil[3]

[1] Dept of Information Technology, NIT Durgapur, Durgapur-713209, India
[2] School of Information Technology, IIT Kharagpur, Kharagpur-721302, India
[3] Dept of Computer Science & Engineering, BESU, Shibpur, India
{nanda.jana,smartyadi88,rajdarbar.r}@gmail.com,
js@cs.becs.ac.in

Abstract. Differential Evolution (DE) algorithm is a real parameter encoded evolutionary algorithm for global optimization. In this paper, Levy distributed DE (LevyDE) has been proposed. The main objective of LevyDE algorithm is to introduce a parameter control mechanism in DE based on levy distribution, a heavy tail distribution, for both the mutation and crossover operations. The main emphasis of this paper is to analyze the behavior and dynamics of the LevyDE and make a comparison with other standard algorithms such as DE/best/1/bin **[1]**, DE/rand/1/bin **[1]** and ACDE **[8]** on basis of CEC'05 benchmark functions.

1 Introduction

The Global optimization problem **[2]** can be formalized as a pair $\langle s,f \rangle$, where $S \subseteq R^D$ and $f : S \to R$ is a D-dimensional real valued function. The problem is to find a point $x^* \in S$ such that $f(x^*) \leq f(x), \forall x \in S$. Here, f does not need to be continuous but to be bounded. There is no exact solution for this problem. Therefore, heuristic search algorithms are used to solve this problem efficiently. The Differential evolution (DE) outperforms the other existing algorithms in robust performance and faster execution to find global optimal solution. The DE **[1]** is a population based stochastic meta-heuristic algorithm for global optimization which is known for its simplicity, effectiveness and robustness. Sometime practical experience shows that it does not perform up to the expectations. The performance of the DE depends on the balance of exploration and exploitation strategies like other evolutionary algorithms. If the balance is hampered, the problem like stagnation of the population, premature convergence etc may appear. The situation when the algorithm does not show any improvement, though it accepts new individuals in the population is known as stagnation. Besides this, premature convergence arises when there is a loss of diversity in the population **[3]**. It generally arises when the objective function is non-separable multimodal having several local and global optimums. To keep the balance between exploration and exploitation strategies, two techniques such as parameter tuning and parameter controlling can be used according to Eiben et al **[4]**. Parameter tuning is the commonly practiced approach that amounts to finding good values for

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 605–613, 2013.

the parameters before the run of the algorithm and then running the algorithm using these values, which remain fixed during the run. This is a trial & error approach. Parameter control is an alternative which controls the parameter in every generation following some specified rules. It has three categories: (a) deterministic parameter control **[5,6]** (b) adaptive parameter control **[7, 8]** and (c) self adaptive parameter control **[9, 10, 11]**.

In this paper, a self adaptive control mechanism is introduced to control the scale parameter and crossover probability using Levy distribution with variable location parameter, because it produces significant amount of changes in the control parameters and reduces manual parameters setting. To validate the proposed strategy, experiments are performed on 10 benchmark test problems that were introduced by Suganthan et al. in CEC 2005 **[2]**. The obtained results are compared against the two conventional DE algorithms **[1]** and Adaptive Differential Evolution Algorithm (ACDE) proposed by Millie Pant et al **[8]**.

2 Related Works

There is a lot of research done on the parameter control mechanisms. The main objective of these researches is to reduce manual control and initialization of the scale factor and crossover probability and to reach the global optimum efficiently. Abbas et al. [14] introduced a self-adaptive approach to DE parameters using variable step length generated by a Gaussian distribution. These parameters are evolved during the optimization process. Liu and Lampinen introduced an adaptive parameter control mechanism using Fuzzy controller in [7]. Yang et al. [10] proposed a self adaptive differential evolution algorithm with neighborhood search (SaNSDE). SaNSDE proposes three self-adaptive strategies: self-adaptive choice of the mutation strategy between two alternatives, self-adaptation of the scale factor F, and self-adaptation of the crossover rate Cr. Qin and Suganthan [11] proposed a self-adaptive DE, called SaDE. In this proposed method, scale factor and mutation probability need not require any predefining. Teo [9] proposed a self-adaptive strategy where population size parameter is adapted during the optimization. A new Differential Evolution algorithm based on Adaptive Control parameters (ACDE) is introduced by Pant et al in [8].

3 Differential Evolution Algorithm

Differential Evolution (DE) is a stochastic, population-based optimization algorithm which was proposed by Storn and Price in 1996 **[1]**. It was developed to optimize real parameter, real valued functions. Global optimization is necessary in fields such as engineering, statistics and finance. But many practical problems have objective functions that are non-differentiable, non-continuous, non-linear, noisy, flat, multi-dimensional or have many local minima, constraints or stochasticity. Such problems are difficult if not impossible to solve analytically. DE can be used to find approximate solutions to such problems. DE is an Evolutionary algorithm whose initial population is of size N_p and dimension is D. The population matrix is initialized as follows:

$$X_i = B_i^L + \text{rand}(0,1) * \left(B_i^U - B_i^L \right) \tag{1}$$

Each initialized vector is called 'target vector'. Here, B_i^U and B_i^L represent upper and lower bound of X_i respectively.

The classical DE has three operators: mutation, crossover and selection.

Mutation: Mutation is a kind of exploration technique that can explore the search space rapidly. It creates donor vectors. Mutation strategies are as follows:

1. DE/rand/1:

$$V_{ij} = X_{r_1,j} + F*\left(X_{r_2,j} - X_{r_3,j}\right);\ r_1 \neq r_2 \neq r_3 \tag{2}$$

2. DE/best/1:

$$V_{ij} = X_{best,j} + F*\left(X_{r_1,j} - X_{r_2,j}\right)\ ;\ r_1 \neq r_2 \tag{3}$$

Here F is a control parameter called scale factor. It controls the speed of convergence towards optimal solutions depending on its value. Range of f is given as [0, 2].

Crossover: In DE, Crossover is an exploration technique. It generates trial vector U_{ij}. There are two types of crossover strategies i.e. binomial crossover and exponential crossover. Generally, binomial crossover is used in DE and it is very similar to uniform crossover in evolutionary algorithms. It is described as follows:

$$U_{ij} = \begin{cases} V_{ij}\ ; \text{if rand}(0,1) \leq CR \vee j=\text{jrand} \\ X_{ij}\ ; \text{otherwise} \end{cases} \tag{4}$$

Here, jrand $\in [1, D]$. The CR is crossover probability. It is defined as $CR \in [0,1]$. This crossover suggests that in trial vector, there must be at least one component from the donor vector.

Selection: Selection is the exploitation process. It is a greedy method. Selection is also termed as mother-child competition. As its name indicates, mother X_i and child U_i compete with each other for survival chance in the next generation. The mother-child competition is done as follows:

$$X_i = \begin{cases} U_i\ ; \text{if } f(U_i) \leq f(X_i) \\ X_i\ ; \text{otherwise} \end{cases} \tag{5}$$

4 Levy Distribution

Levy distribution **[12]** is a stable continuous probability distribution for non-negative random variable. It is stable because it has the property that a linear combination of independent copies of the variable has the same distribution, up to location and scale parameters. The Levy probability distribution function is

$$f(x;\alpha,\beta,c,\mu) = \sqrt{\frac{c}{2\pi}} \left(e^{\frac{-c}{2(x-\mu)}} \Big/ (x-\mu)^{\frac{3}{2}} \right) \tag{6}$$

Here, α is the characteristic exponent; β is the skewness parameter which represents the measure of asymmetry; c is the scale parameter which measures the width of the

distribution; μ is the shift or location parameter. In Levy distribution, the value of the parameters are : α=0.5 ; β=1; c>0; $\mu \geq 0$; Like all stable distributions except the normal distribution, the wing of the probability density function exhibits heavy tail behavior falling off according to a power law :

$$\lim_{x\to\infty} f(x; 0.5,1,c,\mu) = \sqrt{\frac{c}{2\pi}} \left(1 \Big/ x^{\frac{3}{2}} \right) \tag{7}$$

5 LevyDE: Levy Distributed Differential Evolution

This paper introduces a new self-adaptive strategy for control parameters in the DE algorithm. Form the study of **[13]**, it has been observed that heavy tail distribution works better to probabilistically control the parameters. To create a considerable amount of diversity in the solution, a new heavy tail distribution is used in the scale factor and crossover rate. Therefore, these values will be altered according to the following Levy distribution described in equation **(6)**. The reason of choosing levy distribution is that the heavy tail will generate considerable changes more frequently and it has higher probability to do a long jump that may escape from local optima or move away from a plateau that might provide a better solution for multimodal optimization problem.

In LevyDE, scale factor, $F \in [0.5 , 0.9]$ and crossover rate, $CR \in [0.5,1]$. This F and CR are updated according to the Levy distribution. Here, two new parameters are introduced delF and delCR. These parameters are used to control the updation of the shift parameters (i.e. scaleF and scaleCR) of the Levy distribution to get updated control parameters. The parameters scaleF, scaleCR $\in$ [0,1], are initialized randomly. The LevyDE algorithm is as follows:

Procedure LevyDE ()

```
1.  Initialize the population matrix X[N, D] randomly in the search
    space by equation 1 and other parameters.
2.  fitness := calculateFitness(X)
3.  while termination condition not satisfied
    3.1 if delF < U(0, 1)
            3.1.1  scaleF := scaleF + (scaleF^u - scaleF^l)*U(0, 1)
            3.1.2  F := F + (F_max - F_min) *
                   Levy(iteration,0.5,1,1,scaleF)
    3.2 else
            3.2.1  F := F + (F_max - F_min) * Levy
                   (iteration,0.5,1,1,scaleF)
    3.3 if F > F_max
            3.3.1  F := F_min + 0.4*U(0, 1)
    3.4 if delCR < U(0, 1)
            3.4.1  scaleCR := scaleCR + (scaleCR^u - scaleCR^l)*U(0, 1)
            3.4.2  CR := CR + CR_max* Levy (iteration,0.5,1,1,scaleCR)
    3.5 else
            3.5.1  CR := CR + (CR_max - CR_min) *
                   Levy(iteration,0.5,1,1,scaleCR)
```

```
3.6 if CR > CR_max
        3.6.1   CR := CR_max*U(0, 1)
3.7 U := mutation(X, type)
3.8 V := crossover(X, U, type)
3.9 fitness_trial := calculateFitness(V)
3.10     for i := 1 to N in the step of 1
        3.10.1 if fitness_trial_i ≤ fitness_i
                3.10.1.1        X_i := V_i
                3.10.1.2    fitness_i := fitness_trial_i
```

6 Experiments and Analysis of the Results

To test the algorithm, we have selected CEC 2005 benchmark problems **[2]** and considered 10 different functions, described in Table 1. The function 1- 4 is unimodal and rest is multimodal functions. The F10 is an expanded multimodal benchmark function.

Table 1. Benchmark functions selected from CEC 2005 function sets

Function Definition	Dimension	Range	Optimum
$f_1(X)=\sum_{i=1}^{D} X_i^2+f_{bias}$; X=X-o ;	10	[-100,100]	- 450
$f_2(X)=\sum_{i=1}^{D}\left(\sum_{j=1}^{i} X_j\right)^2+f_{bias}$; X=X-o	10	[-100,100]	- 450
$f_3(X)=(A_i X-B_i)+f_{bias}$; $\det(A)\neq 0, a_{ij}\in[-500,500], B_i=A_i * o, o_i\in[-100,100]$	10	[-100,100]	- 310
$f_4(X)=\sum_{i=1}^{D}\left(100\left(X_i^2-X_{i+1}\right)^2+(X_i-1)^2\right)+f_{bias}$; X=(X-o + 1)	10	[-100,100]	390
$f_5(X)=-20e^{\left(-0.2\sqrt{\left(\sum_{i=1}^{D} X_i^2/D\right)}\right)}-e^{\left(\sum_{i=1}^{D}\cos(2\pi X_i)/D\right)}+20+e+f_{bias}$; X=(X-o)*M	10	[-32,32]	-140
$f_6(X)=\sum_{i=1}^{D}\left(X_i^2-10\cos(2\pi X_i)+10\right)+f_{bias}$; X=(X-o)	10	[-5,5]	-330
$f_7(X)=\sum_{i=1}^{D}\left(X_i^2-10\cos(2\pi X_i)+10\right)+f_{bias}$; X=(X-o)*M	10	[-5,5]	-330
$f_8(X)=\sum_{i=1}^{D}\left(\sum_{k=0}^{k_{max}}\left[a_k\cos(2\pi b_k(X_i+0.5))\right]\right)-D\sum_{k=0}^{k_{max}}\left[a_k\cos(2\pi b_k\times 0.5)\right]+f_{bias}$; a=0.5, b=3, k_{max}=20; X=(X-o)*M	10	[-0.5,0.5]	90
$f_9(X)=\sum_{i=1}^{D-1} F_G(F_R(X_i,X_{i-1}))+F_G(F_R(X_D,X_1))+f_{bias}$; $F_G=\sum_{i=1}^{D} X_i^2/4000-\prod_{i=1}^{D}\cos\left(X_i/\sqrt{i}\right)+1$, $F_G=\sum_{i=1}^{D}\left(100\left(X_i^2-X_{i+1}\right)^2+(X_i-1)^2\right)$;X=X-o+1	10	[-π, π]	-460
$f_{10}(X)=\sum_{i=1}^{D-1} G(X_i,X_{i+1})+G(X_D,X_1)+f_{bias}$;$G(X,Y)=0.5+\left(\sin^2\sqrt{(X^2+Y^2)}-0.5\right)/\left(1+0.001(X^2+Y^2)\right)^2$;X=(X-o)*M	10	[-100,100]	-300

The algorithms are tested on Intel® Pentium(R) CPU B960 @ 2.20GHz × 2 processor with 4 GB DDR3 RAM. The operating system platform is Ubuntu 12.04, 32 bit. The programming language is MATLAB R2012a.

For basic DE, F=0.5 and CR=0.50. For LevyDE, F ∈ [0.5, 0.9], CR ∈ [0.5, 1.0], delF= 0.50 and delCR=0.70. The maximum iteration is fixed at D*1e+04 where D is the dimension of the problem. Here D is 10 and population size is 20. In our experiment, four algorithms i.e. LevyDE, DE/best/1/bin, DE/rand/1/bin and Adaptive Differential Evolution algorithm (ACDE) **[8]** are run for 25 times for each of the functions. The results are described in Table 2. The Best Result, Mean Error, Standard Deviation and Success Rate are considered for comparison of different algorithms. We have considered error value 1e-06 for function 1-3, 1e-02 for others.

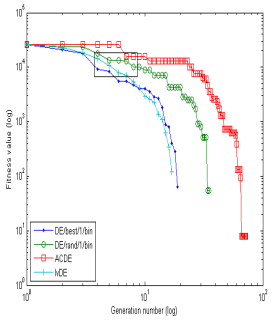


Fig. 1. Convergence plot of 'F1'

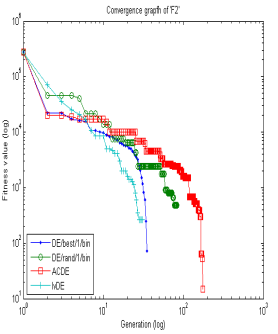


Fig. 2. Convergence plot of 'F2'

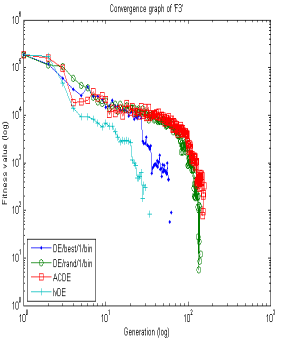


Fig. 3. Convergence plot of 'F3'

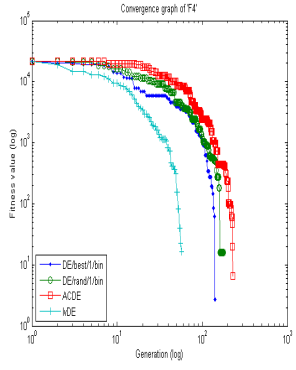


Fig. 4. Convergence plot of 'F4'

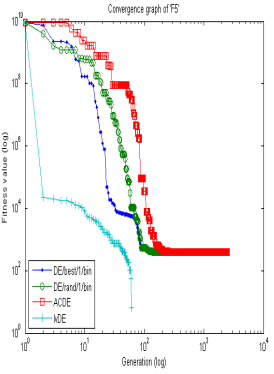


Fig. 5. Convergence plot of 'F5'

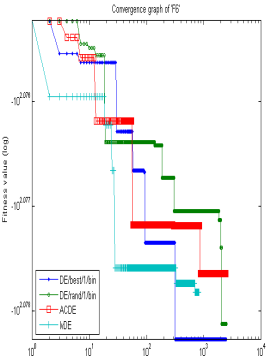


Fig. 6. Convergence plot of 'F6'

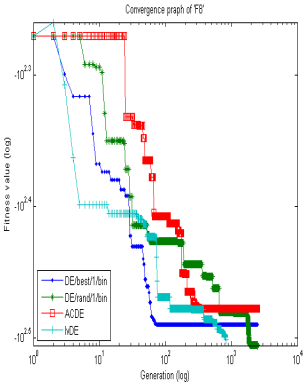


Fig. 7. Convergence plot of 'F8'

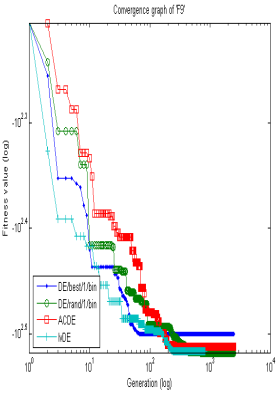


Fig. 8. Convergence plot of 'F9'

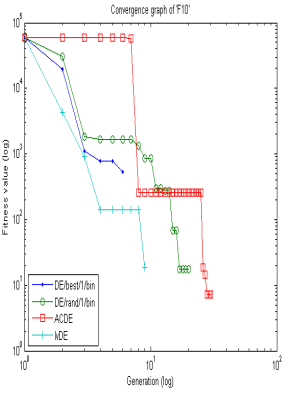


Fig. 1. Convergence plot of 'F10'

Table 2. Comparison analysis of the best result, mean error, standard deviation and success rate of DE/best/1/bin, DE/rand/1/bin, ACDE, and LevyDE

Functions	Procedures	Best result	Mean Error	Std Deviation	Success rate (%)
F1	DE/best	-4.500000e+02	2.042965e-07	1.670335e-07	100
	DE/rand	-4.500000e+02	1.561729e-07	1.151371e-07	100
	ACDE	-4.500000e+02	4.815826e-08	1.358376e-07	100
	LevyDE	**-4.500000e+02**	**2.004074e-09**	**1.065551e-07**	**100**
F2	DE/best	-4.500000e+02	2.695379e-07	2.090216e-07	100
	DE/rand	-4.500000e+02	1.717866e-02	8.583652e-02	084
	ACDE	-4.499767e+02	2.325408e-02	1.867916e-07	000
	LevyDE	**-4.500000e+02**	**1.732750e-07**	**1.260445e-07**	**100**
F3	DE/best	-4.500000e+02	2.611088e-07	1.922718e-07	100
	DE/rand	-4.500000e+02	1.104072e-07	1.096909e-07	100
	ACDE	-4.499985e+02	1.338269e+00	2.899092e+00	000
	LevyDE	**-4.500000e+02**	**1.002874e-07**	**1.809248e-07**	**100**
F4	DE/best	-3.100000e+02	**8.984726e-08**	**5.933756e-08**	100
	DE/rand	-3.100000e+02	9.914377e-02	4.593829e-01	076
	ACDE	-3.061221e+02	3.877860e+00	1.668932e+00	000
	LevyDE	**-3.100000e+02**	7.7675e-07	1.944300e-07	**100**
F5	DE/best	3.900060e+02	4.583631e+00	1.356022e+01	024
	DE/rand	3.900059e+02	4.339623e+00	3.108901e-01	012
	ACDE	3.966768e+02	6.666787e+00	5.933756e+00	000
	LevyDE	**3.900011e+02**	**1.255594e+00**	**2.189511e+00**	**072**
F6	DE/best	-1.197357e+02	2.036411e+01	**6.547529e-02**	000
	DE/rand	-1.198571e+02	2.038093e+01	8.970581e-02	000
	ACDE	-1.198655e+02	2.038609e+01	9.796716e-02	000
	LevyDE	**-1.208680e+02**	**2.036857e+01**	9.175338e-02	000
F7	DE/best	-3.270151e+02	8.745213e+00	5.989216e+00	000
	DE/rand	**-3.300943e+02**	**2.584488e+00**	**3.332907e+00**	**036**
	ACDE	-3.290050e+02	8.996444e+00	5.631635e+00	000
	LevyDE	-3.290050e+02	1.643564e+01	3.854307e+00	000
F8	DE/best	-3.270151e+02	**1.460600e+01**	7.608035e+00	000
	DE/rand	-3.260202e+02	1.701437e+01	**7.496491e+00**	000
	ACDE	-3.190555e+02	2.452018e+01	1.306585e+01	000

Table 2.(*Continued*)

	LevyDE	**-3.290050e+02**	1.643564e+01	1.078447e+01	000
F9	DE/best	1.120927e+02	2.208274e+01	1.450389e-14	000
	DE/rand	1.120927e+02	2.208274e+01	1.450389e-14	000
	ACDE	1.120927e+02	2.208274e+01	1.450389e-14	000
	LevyDE	**1.012927e+02**	**2.208274e+01**	**1.450389e-14**	000
F10	DE/best	-1.299605e+02	8.421141e-01	9.943467e-01	000
	DE/rand	-1.295507e+02	8.334355e-01	**1.997779e-01**	000
	ACDE	-1.299913e+02	**5.945683e-01**	1.792737e+00	000
	LevyDE	**-1.299208e+02**	8.119437e-01	3.954782e-01	000

From the results of Table 2, it is prominent that the LevyDE outperforms the rest 3 algorithms for most of these benchmark functions in all four criteria. The following Figure 1 – 9 shows the convergence rate plot of the four algorithms for the above mentioned function values with respect to generations in the log scale. These figures also show that the convergence rate of LevyDE is better than the rest three algorithms.

7 Conclusion

In this paper, we proposed a modified version of basic Differential Evolution (DE) using Levy distribution for parameter adaption. Here, we have applied 'DE/best/1' mutation strategy and binary crossover strategy with parameter adaption by Levy distribution to improve the exploration and exploitation mechanism of mutation and selection procedure. This mutation strategy improves the problem of premature convergence and trapping in local optimal. Experimental results on standard benchmark functions prove the better efficiency of LevyDE algorithm. Future work is to evaluate the performance of LevyDE using more test functions and make it more robust by upgrading and balancing the exploration and exploitation strategies.

References

[1] Price, K., Storn, R., Lampinen, J.: Differential Evolution - A Practical Approach to Global optimization. Springer (2005)

[2] Suganthan, P.N., Hansen, N., Liang, J.J., Deb, K., Chen, Y.-P., Auger, A., Tiwari, S.: Problem definitions and evaluation criteria for the CEC05 special session on real-parameter optimization. Technical Report, Nanyang Teechnological University, Singapore (May 2005)

[3] Epitropakis, M.G., Plagianakos, V.P., Vrahatis, M.N.: Balancing the exploration and exploitation capabilities of the Differential Evolution algorithm. In: Proceedings of 2008 IEEE Congress on Evolutionary Computation (CEC 2008), pp. 2686–2693 (2008)

[4] Eiben, A.E., Hinterding, R., Michalewicz, Z.: Parameter Control in Evolutionary Algorithms. IEEE Transaction on Evolutionary Computation 3(2), 124–141 (1999)
[5] Krink, T., Filipič, B., Fogel, G.B.: Noisy optimization problems: A particular challenge for differential evolution. In: Proc. IEEE Congr. Evol. Comput., pp. 332–339 (2004)
[6] Liu, B., Zhang, X., Ma, H.: Hybrid differential evolution for noisy optimization. In: Proc. IEEE Congr. Evol. Comput., pp. 587–592 (June 2008)
[7] Liu, J., Lampinen, J.: A fuzzy adaptive differential evolution algorithm. Soft Comput. 2005 9(6), 448–462 (2005)
[8] Thangaraj, R., Pant, M., Abraham, A.: A Simple Adaptive Differential Evolution Algorithm. In: Proc. NaBIC 2009, pp. 457–462 (2009)
[9] Teo, J.: Exploring Dynamic Self-adaptive Populations in Differential Evolution. Soft Computing - A Fusion of Foundations' Methodologies and Applications 10(8), 673–686 (2006)
[10] Yang, Z., Yao, K.T.X.: Self-adaptive Differential Evolution with Neighborhood Search. In: IEEE World Congress on Computational Intelligence, pp. 1110–1116 (2008)
[11] Qin, A.K., Suganthan, P.N.: Self-adaptive Differential Evolution Algorithm for Numerical Optimization. In: Proc. IEEE Congress on Evolutionary Computation (September 2005)
[12] Applebaum, D.: Lectures on Lévy processes and Stochastic calculus, Braunschweig; Lecture 2: Lévy processes, pp. 37–53. University of Sheffield
[13] Pant, M., Thangaraj, R., Abraham, A., Grosan, C.: Differential Evolution with Laplace Mutation Operator. In: CEC 2009, pp. 2841–2849 (2009)
[14] Abbass, H.A.: The self-adaptive pareto differential evolution algorithm. In: Proc. of 2002 Congress on Evolutionary Computation, pp. 831–836 (2002)

Particle Swarm Optimization with Exploratory Move

Nanda Dulal Jana[1] and Jaya Sil[2]

[1] Department of Information Technology, National Institute of Technology, Durgapur, WB, India
nanda.jana@gmail.com
[2] Department of Computer Science & Technology, Bengal Engineering & Science University, Shibpur, WB, India
js@cs.becs.ac.in

Abstract. Particle Swarm Optimization (PSO) algorithm is a swarm based algorithm deliver good performance in many optimization problems. However, PSO has tendency of trapping into local optima. In the paper, an improved PSO algorithm has been proposed by employing Exploratory Move on global best particle of the swarm called as PSO with exploratory move (ExPSO) algorithm. In the proposed approach in order to preventing PSO algorithm from trapping into local optima, particles are jumped to an unknown position made by the exploratory move. The performance of the ExPSO algorithm has been investigated on a set of eight standard benchmark functions and results are compared with the simple PSO, constriction factor PSO (CFPSO) and inertia weight PSO (IWPSO). The numerical results show that the ExPSO algorithm performs better, robust and statistically significant on most of the test cases.

Keywords: Particle Swarm Optimization, Exploratory Move, Exploration and Exploitation, Local Optima.

1 Introduction

PSO is a population based stochastic optimization algorithm inspired by the social behaviour of bird flocking, firstly introduced by Kennedy and Eberhart [1,2]. In the early stage of development of PSO algorithm, it was used for continuous optimization problems. Now PSO has received more and more attentions by many researchers due to its promising optimization capacity in various fields [3,4]. However, PSO is often trapped into local optima due to premature convergence while the convergence rate decreases in the latter period of evolution. Therefore, accelerating convergence rate and avoiding local optima become the two most important and appealing challenges in PSO research.

To overcome the above drawbacks of PSO, existing methods attempted to improve the performance by introducing variable parameters [5] or modifying the updating equations [6] or adopting the operators of the optimization algorithm

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 614–621, 2013.

[7,8]. In [9], a constriction factor algorithm was introduced to ensure the convergence of PSO. The lack of population diversity in PSO algorithms is the reason of premature convergence [10]. Therefore, in addition of search/move operator to PSO, enhancement of global search capacity is most important to improve its performance. In the paper, an improved PSO is proposed by introducing exploratory move on global best particle of the swarm. In the exploratory move, the current point is perturbed one at a time along each variable in positive and negative direction and the best point is recorded. Therefore, the proposed method creates set of search directions iteratively in such a way so that the search directions completely cover the search space. As a result, global search capability increases for its long jump ability. The proposed method shows the fast convergence speed and greatly overcome the tendency of trapping into local optima.

The rest of the paper is organized as follows: section 2 describes the basic concepts of the PSO. In section 3, exploratory move and the proposed ExPSO algorithm has been described while in section 4 experimental analyses, results and discussion are presented. Section 5 concludes the paper with a possible direction in future works.

2 Particle Swarm Optimization Algorithm

In PSO algorithm, each member of the swarm is called a 'particle' and each particle flies around in the D-dimension search space with a velocity. Each particle in the PSO has a position and a velocity, its evaluation is achieved using the objective function/fitness function (f) of the optimization problem, whose variables are the particle position dimensions. The particle updating method tries to move particles to better positions by accelerating them towards personal best position of a particle and global best position of particles.

In general, a particle moves in a D-dimensional search space and a swarm contains N such particles. The position vector and velocity vector of the i^{th} particle is represented by $X_i = (x_{i1}, x_{i2}, .., x_{iD})$ and $V_i = (v_{i1}, v_{i2}, ..., v_{iD})$ respectively. Each particle maintains a memory of its previous best position which is represented by $X_{pbest} = (x_{pbest1}, x_{pbest2}, ..., x_{pbestD})$ and best of all the particles in the swarm by $X_{gbest} = (x_{gbest1}, x_{gbest2}, .., x_{gbestD})$. The basic PSO algorithm can be described using equations (1)and (2).

$$V_i(t+1) = \omega * V_i(t) + c_1 r_1 (X_{pbesti}(t) - X_i(t)) + c_2 r_2 (X_{gbesti}(t) - X_i(t)) \quad (1)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (2)$$

Where c_1 and c_2 are positive constants and called the acceleration coefficients, r_1 and r_2 are two uniformly distributed random number in the interval [0, 1]. V_i is the velocity of i^{th} individual of dimension D and X_i is the current position of i^{th} individual on dimension D. X_{pbesti} is the best position of the i^{th} particle and X_{gbesti} represents the best position found so far by all particles in the swarm at time t. In Eq. (1), ω is the inertia weight which provides the necessary diversity to the swarm by changing the momentum of particles.

Algorithm 1. Exploratory Move

Input: Initialize current solution (X^c) with dimension D. X_i^c is the i^{th} dimension perturbed byΔ_i
Output: New best point
begin
  for $i = 1$ *to* D **do**
    $y_i^+ = x_i^c + \Delta_i$
    // y_i^+ is the i^{th} component of $Y^+ = (x_1^c, x_2^c, ..., x_i^c + \Delta_i, x_{i+1}^c, ...x_D^c)$
    $y_i^- = x_i^c - \Delta_i$
    // y_i^- is the i^{th} component of $Y^- = (x_1^c, x_2^c, ..., x_i^c - \Delta_i, x_{i+1}^c, ...x_D^c)$
    evaluate $f = f(X^c), f^+ = f^+(Y^+) and f^- = f^-(Y^-)$
    find $f_{min}, min(f, f^+, f^-)$. Set X corresponds to f_{min}
    if $i = D$ **then**
      Break
  if $X^c \neq X$ **then**
    Success
  else
    failure

3 Proposed Methodology

3.1 Exploratory Move

An exploratory move is performed systematically in the vicinity of the current point to find the best point around the current point. In this move, the current point is perturbed in positive and negative directions along each variable one at a time and the best point is recorded. The current point is changed to the best point at the end of each variable perturbation. If the point found at the end of all variable perturbations is different than the original point, the exploratory move becomes success otherwise not. The exploratory move procedure shown in Algorithm 1. In the proposed method, a set of search directions are iteratively generated to cover the search space completely. The process starts from any point in the search space and the search can be reached to any other point in the search space by travelling along the search directions. In an N-dimensional problem, its requires at least N linearly independent search directions. Among many possible combinations of N search directions, some combinations may help to reach the destination faster with less number of iterations.

3.2 Proposed Mechanism

PSO has been shown rapid convergence in the first part of the search and then slow down or no improvement has been observed in the fitness function. This behaviour has been attributed as the loss of diversity in the population. It is possible to lead the swarm away from a current location by improving a single individual if the improved individual becomes the new global best. The global best individual attracts all members of the swarm. In the paper, we consider *gbest* position as the current position and exploratory move is applied on this current position along each component,selected randomly. Suppose

Algorithm 2. Algorithm for Proposed Mechanism

```
Input: Initialize the particles position and velocity
Output: Global best
begin
    Calculate fitness. Updating pbest and gbest
    while (stopping condition is not reached) do
        for i = to N do
            Updating velocity and position using Eq. (1) and (2)
            calculate fitness. updating the pbest and the gbest
        call "Exploratory Move" on global best and find the new global best.
        if the new global best is better than the gobal best, then replace the global best.
```

$X_{gbest} = (x_{gbest1}, x_{gbest2}, ..., x_{gbestD})$ and select a random integer k between 1 to D. Exploratory move is therefore, performed on k^{th} component, X_{gbestk} of X_{gbest} to compute $X_{gbest}(+)$ and $X_{gbest}(-)$ along the positive and negative direction in the search space by random increment and decrement. This is represented by
$X_{gbest}(+) = (x_{gbest1}, x_{gbest2}, ..., x_{gbestk} + r_1, ..., x_{gbestD})$
$X_{gbest}(-) = (x_{gbest1}, x_{gbest2}, ..., x_{gbestk} - r_2, ..., x_{gbestD})$

Evaluating $f(X_{gbest}(+))$ and $f(_{Xgbest}(-))$,the best position is selected based on the fitness values of the points X_{gbest}, $X_{gbest}(+)$and $X_{gbest}(-)$. This point corresponds to the current position of the next iteration of the exploratory move. Exploratory move terminates when fitness value of the best position reaches to the desired error. Otherwise, it is performed on all components of the dimensions. The proposed algorithm is shown in Algorithm 2. The proposed approach escapes local optima and new positions become the new global best position.

4 Experimental Studies

4.1 Benchmark Functions

Eight benchmark functions are considered for experiment, which are widely adopted in global optimization algorithms [11,12]. The test functions f_1 and f_2 are unimodal, having only one global minimum 0. The benchmark functions from f_3 to f_7 are multimodal functions having the global minimum at the origin or very near to the origin. We have also taken a noisy function f_8, where a uniformly distributed random noise is added to the function. The description of these benchmark functions and their global optima are given in Table 1.

4.2 Parameter Settings

For the purpose of performance evaluation, we compare the purposed algorithm with other PSO algorithms, simple PSO, constriction factors PSO (CFPSO) and inertia weights PSO (IWPSO) over 50 independent runs. Experiment is carried out for eight benchmark problems having 30 dimensions and population size is 50. The parameters of the proposed algorithm are $c_1 = c_2 = 1.49618$ and decreasing inertia weight (ω) in each iteration starting from 0.9 to 0.4. The acceleration

Table 1. The Benchmark Functions

Function	Mathematical Representation	Range	Option
Sphere	$f_1(x) = \sum_{i=1}^{D} x_i^2$	$[-100, 100]$	0
Schwefel	$f_2(x) = \sum_{i=1}^{D} \sum_{j=1}^{j} x_i^2$	$[-100, 100]$	0
Griewank	$f_3(x) = \frac{1}{40000} \sum_{i=1}^{D} x_i^2 - \prod_{i=1}^{D} cos(\frac{x_i}{\sqrt{i}}) + 1$	$[-600, 600]$	0
Rastrigin	$f_4(x) = \sum_{i=1}^{D} (x_i^2 - 10cos(2\pi x_i) + 10)$	$[-5.12, 5.12]$	0
Rosenbrock	$f_5(x) = \sum_{i=1}^{D-1} [(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2]$	$[-100, 100]$	0
Ackley	$f_6(x) = 20 + e - 20e^{-\frac{1}{5}\sqrt{\frac{1}{n}\sum_{i=1}^{D} x_i^2}}$ $-e^{\frac{1}{n}\sum_{i=1}^{D} cos(2\pi x_i)}$	$[-32, 32]$	0
Weierstrass	$f_7(x) = \sum_{i=1}^{D} (\sum_{k=0}^{k_{max}} [a^k cos(2\pi b^k (x_i + 0.5))])$ $-n \sum_{k=0}^{k_{max}} [a^k cos(2\pi b^k 0.5)]$ with $a = 0.5$, $b = 3$ and $k_{max} = 20$	$[-100, 100]$	0
Dejong's Noisy Function	$f_8(x) = \sum_{i=0}^{D-1} (i + 1)x_i^4 + rand[0, 1]$	$[-1.28, 1.28]$	0

coefficients ($c_1 and c_2$) for the simple PSO, CFPSO and IWPSO are set to 2. The inertia weight (ω) for simple PSO is set to 0.732 and for IWPSO is in decreasing order and in each iterations set from 0.9 to 0.4. Maximum velocity, $V_{max} = X_{max}$ where $[X_{min}, X_{max}]$ is the search space range. The same initial population is used for all PSO algorithms. In this work, the termination criteria are considered as maximum number of generations i.e. 4000 and $E = |f(X) - f(X^*)| \leq e$ ($f(X)$ is the current best and $f(X^*)$is the global optimum) is the best-error of a run of the algorithm and e is the threshold error. In our experiment error $e = 0.001$. Algorithms are implemented using MATLAB 7.6.0 (R2008a) applied on Intel (R) Core (TM) i7-2670QM CPU @ 2.20 GHz with 8 GB RAM on windows 7 Home Premium platform.

4.3 Results and Discussion

Table 2 presents the mean, standard deviation, average number of generations and success rate (frequency of hitting the optimum) of the benchmark functions respectively using the four PSO algorithms over 50 independent runs respectively. The best results are marked in boldface. Convergence characteristics of each functions are comaperd with the PSO algorithms are shown in Figure 1. From the results it has been observed that the proposed method outperforms over other algorithms with 100% Success Rate(SR) for the function f_1, f_2, f_4, f_6 and f_7. In case of the functions f_3 and f_5, proposed ExPSO algorithm produced better solution but no improvement in SR as well as convergence speed. The ExPSO achieved better results than other algorithms with zero SR for function f_8. The quality of solution obtained by the proposed method with minimum number of average generations than simple PSO, CFPSO and IWPSO except the function f_8 as shown in Table2.

Table 3 shows results of unpaired t-tests between the best algorithm and the second best in each case (standard error difference of the two means, 95% confidence interval of this difference, the t- value and the two - tailed P value). For all cases in Table 3 sample size is 50 and degree of freedom is 98. It is interesting

Table 2. Result of 8 functions

Function no.	Algorithm	Evaluation Metrics			
		Mean	Std. Dev.	Avg no. Generations	SR
f_1	ExPOS	**8.740e-004**	**1.343e-004**	**165.36**	**100.00**
	PSO	$2.391e+001$	$1.766e+001$	4000.00	0.00
	CFPSO	$2.431e+003$	$2.314e+002$	4000.00	0.00
	IWPSO	$5.200e+003$	$6.141e+003$	3438.36	54.00
f_2	ExPOS	**9.864e-004**	**1.436e-005**	**1944.22**	**100.00**
	PSO	$1.618e+004$	$5.285e+003$	4000.00	0.00
	CFPSO	$4.612e+003$	$7.257e+002$	4000.00	0.00
	IWPSO	$5.804e+004$	$2.953e+004$	4000.00	0.00
f_3	ExPOS	**1.259e-001**	**8.128e-002**	**3870.20**	**4.00**
	PSO	$3.012e+000$	$1.270e+001$	4000.00	0.00
	CFPSO	$2.294e+001$	$2.155e+000$	4000.00	0.00
	IWPSO	$4.340e+001$	$5.548e+001$	3758.00	24.00
f_4	ExPOS	**9.553e-004**	**4.573e-005**	**767.08**	**100.00**
	PSO	$1.020e+002$	$2.627e+001$	4000.00	0.00
	CFPSO	$1.882e+002$	$1.227e+001$	4000.00	0.00
	IWPSO	$1.697e+002$	$4.690e+001$	4000.00	0.00
f_5	ExPOS	**1.280e+001**	**2.740e+001**	**3468.58**	**30.00**
	PSO	$6.024e+005$	$5.174e+005$	4000.00	0.00
	CFPSO	$4.508e+007$	$1.010e+007$	4000.00	0.00
	IWPSO	$7.600e+005$	$4.314e+005$	4000.00	0.00
f_6	ExPOS	**9.571e-004**	**4.182e-005**	**834.34**	**100.00**
	PSO	$3.216e+000$	$1.918e+000$	4000.00	0.00
	CFPSO	$1.041e+001$	$3.605e-001$	4000.00	0.00
	IWPSO	$1.996e+001$	$3.048e-003$	4000.00	0.00
f_7	ExPOS	**0.000e+000**	**0.000e+000**	**3.96**	**100.00**
	PSO	$1.819e-005$	$1.286e-004$	24.02	**100.00**
	CFPSO	$1.053e+001$	$1.124e+001$	3714.36	14.00
	IWPSO	$1.383e-013$	$2.766e-014$	8.54	**100.00**
f_8	ExPOS	**2.609e-003**	**1.052e-003**	4000.00	0.00
	PSO	$4.004e+003$	$9.258e+003$	4000.00	0.00
	CFPSO	$6.813e+002$	$1.441e+002$	4000.00	0.00
	IWPSO	$5.600e+003$	$1.280e+004$	4000.00	0.00

Table 3. Results of Unpaired t-test on the data of Table 2

Function	Std. Error	T	95% Conf. Interval	Two-Tailed P	Significance
f_1	2.498	9.5732	$(-28.865, -18.953)$	< 0.0001	Extremely Significant
f_2	102.629	44.9383	$(-4815.664, -4408.3341)$	< 0.0001	Extremely Significant
f_3	1.796	1.6074	$(6.451, 0.677)$	0.1112	Not Significant
f_4	3.715	27.4550	$(-109.377, -94.627)$	< 0.0001	Extremely Significant
f_5	73171.410	8.2325	$(-747593.512, -457180.888)$	< 0.0001	Extremely Significant
f_6	0.271	11.8529	$(-3.753, -2.677)$	< 0.0001	Extremely Significant
f_7	0.000	1.0002	$(-0.0001, -0.0000)$	0.3197	Not Significant
f_8	20.379	33.4316	$(-721.739, -640.856)$	< 0.0001	Extremely Significant

to see from Table 2 and Table 3 that in most of the cases the proposed method meets or beats the nearest competitor in a satistically meaningfull way. These results show that the proposed method leads to significant improvements in most cases.

5 Conclusions

The proposed algorithm shows better performance both in early period of generations and later period of generations. Thus it achieves effective trade-off

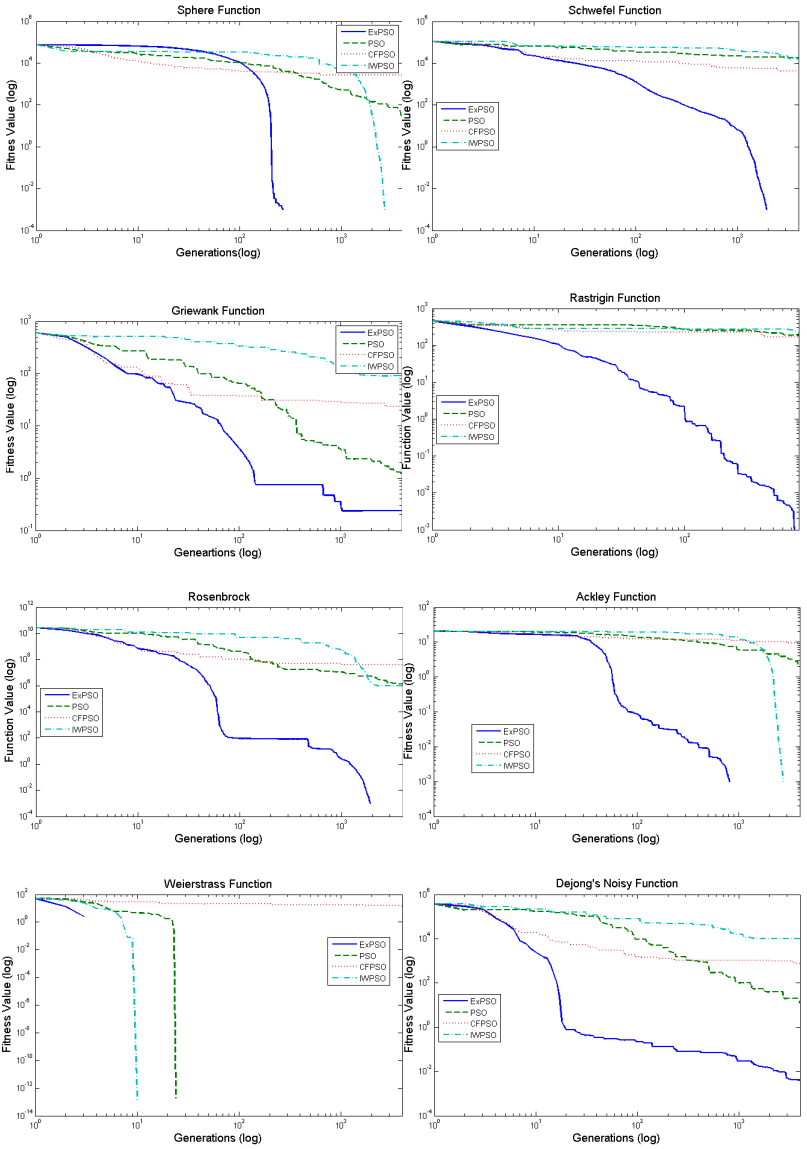


Fig. 1. Convergence characteristics of 8 benchmark functions for 30 dimensions

between exploration and exploitation. The experimental results show that for the optimization problems described by the benchmark functions, our algorithm can obtain better performances than simple PSO, constriction factor PSO and inertia weight PSO. In our future study, a method that can adaptively tune the parameters (inertia weight and constriction factor) of PSO will be investigated using exploratory move.

References

1. Kennedy, J., Eberhart, R.: Particle Swarm Optimization. In: IEEE International Conference on Neural Networks (ICNN 1995), pp. 1942–1948. IEEE Press, Australia (1995)
2. Eberhart, R.C., Kennedy, J.: A New Optimizer Using Particle Swarm Theory. In: 6th International Symposium on Micro Machine and Human Science (ISMMHS 1995), Nagoya, Japan, pp. 39–43 (1995)
3. Ho, S.Y., Lin, H.S., Liauh, W.H., Ho, S.J.: OPSO: Orthogonal Particle Swarm Optimization and its application to task assignment problems. Journal of IEEE Trans. Syst., Man, Cybern. A, Syst., Humans 38(2), 288–298 (2008)
4. Liu, B., Wang, L., Jin, Y.H.: An effective PSO-based memetic algorithm for flow shop scheduling. Journal of IEEE Trans. Syst., Man, Cybern. B, Cybern. 37(1), 18–27 (2007)
5. Jiao, B., Lian, Z.G., Gu, X.S.: A dynamic inertia weight particle swarm optimization algorithm. Journal of Chaos, Solitons and Fractal 37(1), 698–705 (2008)
6. Ling, S.H., Iu, H.H.C., Chan, K.Y.: Hybrid particle swarm optimization with wavelet mutation and its industrial applications. Journal of IEEE Trans. Syst. Man Cybernetics 38(3), 743–763 (2008)
7. Omran, M.G.H., Engelbrecht, A.P., Salman, A.: Differential evolution based particle swarm optimization. In: IEEE Conf. Swarm Intelligence Symposium (SIS), pp. 112–119 (2007)
8. Chen, M.R., Li, X., Zhang, X., Lu, Y.Z.: A novel particle swarm optimizer hybridized with extremal optimization. Journal of Appl. Soft Comput. 10(2), 367–373 (2010)
9. Clerc, M.: The swarm and the queen: towards a deterministic and adaptive particle swarm optimization. In: The Congress on Evolutionary Computation (CEC 1999), pp. 1951–1957. IEEE Service Center, Piscataway (1999)
10. Ratnaweera, A., Halgamuge, S.K., Watson, H.C.: Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. Journal of IEEE Transactions on Evolutionary Computation 8(3), 240–255 (2004)
11. Liang, J.J., Qin, A.K., Suganthan, P.N., Baskar, S.: Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. Journal of IEEE Trans. Evol. Comput. 10(3), 281–295 (2006)
12. Zhan, Z., Zhang, J., Li, Y., Chung, H.S.: Adaptive Particle Swarm Optimization. Journal of IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 39(6), 1362–1381 (2009)

Multi Objective Optimization of Expense and Revenue in a Cognitive Radio Network Using NSGA-II

Subhasree Bhattacharjee[1], Suman Bhattacharjee[2], and Roukna Sengupta[3]

[1] Dept. of CA, Narula Institute of Technology, Kolkata, India
bhattacharjeesubhasree@gmail.com
[2] IBM INDIA Pvt. Ltd. kolkata
sumanbhattacharjee@in.ibm.com
[3] Dept. of CSE, RCC Institute of Information Technology, Kolkata, India
rouknasengupta@gmail.com

Abstract. In this paper we consider an auction framework of cognitive radio network comprises of primary and secondary users (SU). The spectrum is divided into channels using frequency division multiple access (FDMA). Primary users have license to use the channels. When the channels are idle, primary users lease the vacant spectrum for monetary gain. Cognitive users or SUs bid for the channels. The purchaser who provides highest bid value is selected by primary user. Our objective is to maximize the revenue earned by primary users and to minimize the expense given by the secondary users. The problem is bi-objective and both the objectives are conflicting. Using Non dominated sorting genetic algorithm II; we solve both the objectives of primary and secondary users. The algorithm solves the problem well and find optimize values of both expense and revenue.

1 Introduction

In the last few decades there is an exponential growth of wireless devices. Now, the key focus of wireless service providers is to increase the efficiency of use of available spectrum (band). The recent trend of explorations is on the cognitive use of radio spectrum which is popularly known in literature as cognitive radio. In the CR context, there are two types of users namely licensed users (primary users) and unlicensed users (also known as cognitive users or secondary users). In this context of CR, by leasing the channels to secondary users, primary users will get incentives in the form of rebate which can be considered as revenue for primary users and this revenue of primary users is directly correlated to the expenses incurred by secondary users for the usage of licensed bands allocated to primary users. So, the allocation problem is here to strike a balance between revenue of primary users and expenses of secondary users. Considering primary users and secondary users are selfish entities. The above mentioned objective of the allocation problem is conflicting in nature. In this paper we focused on formulating the problem as an optimization problem and solve it by using NSGA-II.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 622–627, 2013.

To the best of our knowledge and as revealed by our literature survey that till date there is no effort made by researchers to view this allocation problem as a conflicting multi objective problem to solve it. However, here we present some of the works related to this domain. In [1] authors propose rules for secondary users to use idle spectrum band but have to ensure that interference is within the threshold limit. In [2], secondary user's spectrum access pattern is followed and a suitable spectrum sharing scheme is proposed. Auction strategy brings a good approach of spectrum sharing, if we consider the economic aspect [3],[4].

Among the various evolutionary algorithms, the NSGA II and SPEA2 are the most popular algorithms. Non dominated sorting Genetic Algorithm (NSGA-II) is a particular evolutionary algorithm that has been used for multi objective optimization.

The rest of the paper is organized as follows. In section 2, construction of objective function for channel allocation is discussed. In section 3, we have introduced the solution of the problem by using NSGA-II. In section 4, experiments and results are discussed. We conclude the paper in section 5.

2 System Model

In this paper we consider multi cell cognitive radio network in which there are multiple primary and secondary users. Numbers of channels for each cell is equal to M. When the channels are vacant, primary users apply marketing strategy for auctioning the channels. Secondary users provide bids for channels. The auctioneer selects the purchaser who provides highest bid value.. The main objective is to make both purchaser and the seller benefitted.

Let, rev be an N X K channel assignment matrix where rev (PU,ch)= rev_{PU}^{ch} = r, if PU earns revenue r for channel ch. 0,if that channel is not assigned to PU. We can formulate the optimization problem of maximizing the revenue earned by all primary users in the system. The optimization problem is of the form:

$$arg\ max \sum_{PU=1}^{N} \sum_{ch=1}^{K} rev_{PU}^{ch} \tag{1}$$

Subject to the conditions that 1. Each primary user requires at most one channel .2. Each channel cannot be used by more than one primary user. The conditions are as follows:

$$c_{\min} \leq \sum_{ch=1}^{K} rev_{PU}^{ch} < c_{\max} \tag{2}$$

$$c_{\min} \leq \sum_{PU=1}^{N} rev_{PU}^{ch} < c_{\max} \tag{3}$$

Here, $c_{\min}$ is equal to the minimum revenue of a channel that a primary user earns and $c_{\max}$ is the maximum revenue of a channel earned by primary user.

The resultant constrained optimization problem is Max z1=

$$\sum_{PU=1}^{N}\sum_{ch=1}^{K} rev_{PU}^{ch} + \lambda_1(c_{\max} - \sum_{ch=1}^{K} rev_{PU}^{ch}) + \lambda_2(c_{\max} - \sum_{PU=1}^{N} rev_{PU}^{ch}) \quad (4)$$

$\lambda 1$, λ_2 are Lagrange coefficients. Whose values vary from +30 to -30.Equation (4) is our first objective function. Now, if expense be an M X K channel assignment matrix where expense (SU,ch)= $expense_{SU}^{ch}$ = e, if SU has expense value e for purchasing channel ch. 0, if that SU will not purchase the channel ch. We can formulate the optimization problem of minimizing the expense provided by all secondary users in the system. The optimization problem is of the form:

$$arg\ min \sum_{SU=1}^{M}\sum_{ch=1}^{K} \exp ense_{SU}^{ch} \quad (5)$$

Subject to the condition that total amount of bid provided by a single secondary user should not exceed a threshold value (b_{max})

$$0 \le \sum_{ch=1}^{K} \exp ense_{SU}^{ch} \le b_{\max} \quad (6)$$

The resultant constrained optimization problem is Min z2=

$$\sum_{SU=1}^{M}\sum_{ch=1}^{K} \exp ense_{SU}^{ch} - \lambda_3(b_{\max} - \sum_{ch=1}^{K} \exp ense_{SU}^{ch}) \quad (7)$$

Where, $\lambda 3$ is Lagrange coefficient. Equation (7) depicts our second objective function.

3 Solving the Problem Using NSGA-II

We firstly describe NSGA-II and then outline the pseudo code for the proposed system.

3.1 Non Dominated Sorting Genetic Algorithm

A Pareto optimal solution set is to be determined for solving multi objective optimization problem [5]. Deb et. al. proposed an evolutionary algorithm which is well known as non dominated sorting GA-II (NSGA-II). NSGA-II is very simple and it has better spread of solutions and better convergence near the Pareto optimal front.

Pseudo code

Input: Initial Revenue matrix and expense matrix. **Output**: Optimum value of revenue and expense for all users and for all channels.

Begin

Set: rev_i^c ←real number, $expense_j^c$ ←real number ,$demand_c$ ← real number ;

Repeat

Call NSGA-II (rev $_i^c$, expense $_j^c$, demand $_c$, mutprob, crprob, Gener, MAXGENER);
Move-to (rev $_i^c$, expense $_j^c$);
Until Gener<=MAXGener
End.
Procedure NSGA-II (rev $_i^c$, expense $_j^c$, demand $_c$, mutprob, crprob, Gener, MAXGENER)
Begin
Initialize a random parent population of size N;
Sort initial population based on the non-domination.
For
K< MAXGENER do
Begin
Create child population (Q_t) using following three
operations: 1) Binary tournament selection, 2)
recombination, and 3) mutation;
Combine parent and child population to form
merged population i.e. $R_t = P_t \cup Q_t$
Construct all non dominated Front sets
($F_1, F_2, \ldots\ldots$);
Set: $P_{t+1} = \emptyset$; i=0
Repeat
$P_{t+1} = P_{t+1} \cup F_i$;
i=i+1;
Until $|P_{t+1}| + |F_i| \leq N$
End.
Calculate crowding distance in F_i ;
Based on the crowding distance sort F_i in descending order;
$P_{t+1} = P_{t+1} \cup$ First ($N - |P_{t+1}|$) elements of F_i ;
End.
End For
End.
Update
1. Expense $_j^c$
2. Demand $_c$ using equation (7)
3. rev $_i^c$

Return
End

4 Result and Discussion

The revenue and expense matrices are initialized randomly. After the convergence of NSGA-II, the best chromosome is obtained from the Pareto front. To select the best chromosome in the Pareto front we propose a metric. This metric is the ratio of normalized value of revenue and normalized value of expense. Let, $f1_i$ (.) and $f2_i$ (.)

are the measures of revenue and expense for i^{th} chromosome and $f1_i*(.)$ and $f2_i*(.)$ are the normalized values of revenue and expense. The normalized values can be defined as

$$f1^*_i(.) = \frac{f1_i}{\sum_{i=1}^{n} f1_i} \quad \text{and} \quad f2^*_i(.) = \frac{f2_i}{\sum_{i=1}^{n} f2_i}$$

The above equation is repeated for all chromosomes in the Pareto front. Next the ratio of normalized value of revenue and normalized value of expense is calculated for all chromosomes in the Pareto front. This new measure is defined as

$$P^*_i(.) = \frac{f1^*_i}{f2^*_i} \text{.....(8)}$$

The effective chromosome $chrom_j$ for j=1 to popsize, is the one which has largest $P^*_j(.)$ value. Figure 1 shows the variation of objective functions with iteration. We vary the number of iteration from 50 to 200.We observe that the first objective function f1 increases as iteration increases and the second objective function f2 decreases as iteration increases. It is seen from the graph that at point z, f1 and f2 intersect each other. At this point f1=f2. At this point revenue earned by primary users=expense paid by secondary users. This point z is termed as 'zero point'. Pareto optimal fronts in the spectrum auction problem are shown in Fig.2 for different generations. The effective solutions on each Pareto front have been marked in figure 2 . The effective solutions indicate the chromosome with maximum normalized ratio of revenue and expense value as expressed in equation (8).

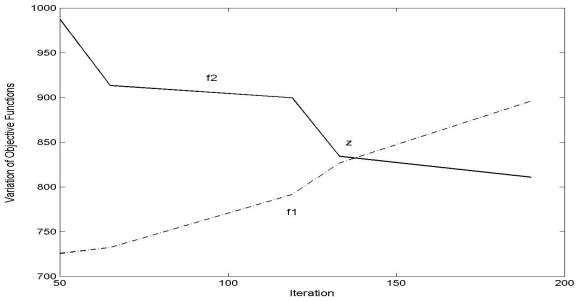


Fig. 1. Variation of objective Functions with Iteration

Table 1. Comparison of our present and previous work

Method Used	Revenue Earned
Simple Greedy Method	887.28
Our Approach using NSGA-II	896.1892

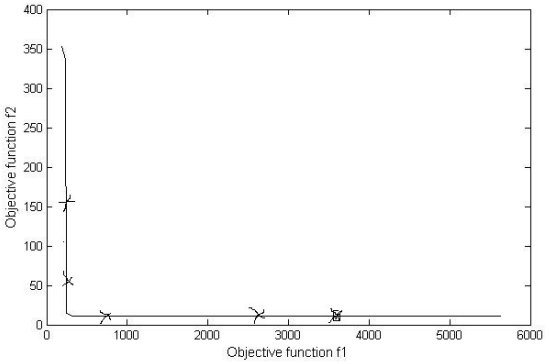


Fig. 2. Non Dominated Solutions for generation up to 75

5 Conclusions

In this paper, we consider a cognitive radio network where primary users perform auction to sell the vacant spectrum to the secondary users. Our objectives are to maximize the revenue earned by primary users and minimize the expense paid by secondary users. We use NSGA-II to optimize the problem. The results are compared with previously used Simple Greedy Method and our approach provides better results.

References

1. Zheng, H., Peng, C.: Collaboration and fairness in opportunistic spectrum access. In: Proc. of IEEE International Conference on Communications (ICC 2005), Seoul, pp. 3132–3136 (May 2005)
2. Keshavamurthy, S., Chandra, K.: Multiplexing analysis for spectrum sharing. In: Proc. of IEEE MILCOMM 2006, Washington D.C., pp. 1–7 (October 2006)
3. Niyato, D., Hossain, E.: Spectrum trading in cognitive radio networks:a market-equilibrium-basedapproach. IEEE Wireless Communications 15(6), 71–80 (2008)
4. Beibei, W., Yongle, W.: Game theoretical mechanism design methods. IEEE Signal Processing Magazine 25(6), 74–84 (2008)
5. Coello, C.A., Pulido, G.T., Lechuga, M.S.: Handling multiple objectives with Particle Swarm Optimization. IEEE Transactions on Evolutionary Computation 3, 256–279 (2004)

Anytime Pack Heuristic Search

Satya Gautam Vadlamudi[1], Sandip Aine[2], and Partha Pratim Chakrabarti[1]

[1] Indian Institute of Technology Kharagpur, Kharagpur, WB 721302, India
{satya,ppchak}@cse.iitkgp.ernet.in
[2] Indraprastha Institute of Information Technology, New Delhi, DL 110020, India
sandip@iiitd.ac.in

Abstract. Heuristic search is a fundamental problem solving technique in artificial intelligence. In this paper, we propose an anytime heuristic search algorithm called Anytime Pack Search (APS) which helps in solving hard combinatorial search problems efficiently. It expands nodes of a search graph in a localized best-first manner so as to converge towards good quality solutions at regular intervals. APS is complete on bounded graphs and guarantees termination with an optimal solution. Experimental results on the sliding-tile puzzle problem, the traveling salesman problem, and the single-machine scheduling problem show that APS significantly outperforms some of the state-of-the-art anytime algorithms.

Keywords: problem solving, heuristic search, anytime algorithms.

1 Introduction

Design of efficient anytime algorithms is one of the key aspects of time limited search based reasoning [2]. Heuristic search based methods are widely used in artificial intelligence for dealing with various hard combinatorial state-space search problems. Since pure best-first search techniques such as A* [3] consume a lot of time and resources for producing solutions, anytime algorithms are developed. Anytime heuristic search algorithms aim at producing good quality solutions quickly and improve upon them as time passes, with the help of heuristic guidance.

Depth-first branch and bound [5] with heuristic guidance can be viewed as a simple anytime heuristic search algorithm. More involved anytime algorithms include: Beam-stack search [10] which makes beam search complete via chronological backtracking and is sometimes referred to as depth-first beam search; Anytime Window A* (AWA*) [1] which uses a window-based depth-guiding mechanism over best-first search to produce solutions in an anytime manner; and Anytime Non-parametric A* (ANA*) [9] which improves upon the parametric weighted-A* based anytime techniques to give rise to a greedy non-parametric version.

We develop a new anytime algorithm inspired by the working of beam search [7] and structural anytime algorithms such as AWA*. The idea is to proceed in a best-first manner over a limited set of nodes– similar to beam search to

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 628–634, 2013.

produce solutions quickly, while ensuring completeness by avoiding inadmissible pruning– similar to structural anytime algorithms. This is achieved by focusing on a selected set of seed nodes in each iteration of the algorithm which we call as *Pack*. The size of the pack can be either given as input or varied using standard mechanisms both of which are demonstrated. Also, the first iteration of the proposed algorithm matches with that of beam search but without the inadmissible pruning. Experimental results indicate that the proposed algorithm gives very good anytime performance compared to the above discussed algorithms.

The rest of the paper is organized as follows: In Section 2, we present the proposed algorithms and their properties. In Section 3, we present the experimental results comparing the proposed algorithms with some of the state-of-the-art anytime heuristic search algorithms, and conclude with the key observations.

2 Proposed Methods

In this section, we present the proposed anytime pack search algorithm and its variations, along with their properties.

Anytime Pack Search (APS) takes as input the search graph and an initial state or start node, similar to other anytime algorithms. Additionally, it takes the *pack size* K as input which is a parameter specific to APS. This parameter can be controlled by the user and it determines the frequency at which solutions are produced by the algorithm. Using a small value of K, one can expect solutions to be reported at a high frequency, and vice-versa.

Algorithm 1. Anytime Pack Search (APS)

1: **INPUT ::** A search graph G, a start node s, and pack size K.
2: $BestSol \leftarrow \infty$; $g(s) \leftarrow 0$; Calculate $f(s)$; $Level(s) \leftarrow 0$;
$SuspendList \leftarrow \{s\}$; $ClosedList \leftarrow \phi$; $OpenList \leftarrow \phi$; $ChildrenList \leftarrow \phi$;
3: **while** $SuspendList \neq \phi$ **do**
4: $\{BestSol, Path\} \leftarrow$ **ExpandKNodes(**$SuspendList, K, BestSol, other\ lists$**)**;
5: **while** $ChildrenList \neq \phi$ **do**
6: Move all nodes from $ChildrenList$ to $OpenList$; (or, equivalently, swap the lists)
$\{BestSol, Path\} \leftarrow$ **ExpandKNodes(**$OpenList, K, BestSol, other\ lists$**)**;
7: **return** $BestSol, Path$;

Algorithm 1 presents the pseudo-code of APS. It uses four lists for its operation: *SuspendList*, *ClosedList*, *OpenList*, and *ChildrenList*. Initially all lists are empty except for *SuspendList* which contains the start node. In each iteration (Lines 4 to 6 of Algorithm 1), initially, K nodes are expanded (the process of generating children) from *SuspendList* which act as a set of seeds for finding a solution in that iteration. The generated children go to *ChildrenList* whose size is capped at K, the remaining of which go to *SuspendList*. Consequently, *OpenList* takes the K nodes from *ChildrenList* for expansion, whose children will go to *ChildrenList* similar to what has been mentioned above. This process will continue until *ChildrenList* is empty which *will* happen since the search is depth-bounded like that of most anytime algorithms. APS terminates when there are no more nodes remaining for expansion. Each iteration can be seen as

an attempt towards producing a better solution by exploring a beam originating from a most promising subset of the unexpanded nodes.

Algorithm 2. ExpandKNodes

```
1: INPUT :: CurList from which K nodes are to be expanded, K, BestSol, and the other lists.
2: for K number of times do
3:    if CurList = φ then
4:       return {BestSol, Path};
5:    n ← least f-valued node from CurList; (for minimization problem)
6:    if IsGoal(n) then
7:       if f(n) < BestSol then
8:          BestSol ← f(n); Path ← Path from s to n obtained by tracing the parent of n;
9:       Move n from CurList to ClosedList; continue;
10:   GenerateChildren(n); Move n from CurList to ClosedList;
11: return {BestSol, Path};
```

Algorithm 2 presents the pseudo-code of the `ExpandKNodes` routine. Most promising node is chosen for expansion each time from the given list which demonstrates the localized best-first nature of APS. If the chosen node is a goal node, the current best solution is updated, otherwise, its children are generated and the node is added to the *ClosedList*.

Algorithm 3. GenerateChildren

```
1: INPUT :: Node n whose children are to be generated, K, and the lists.
2: if Level(n) = MAX_DEPTH − 1 then
3:    return;
4: for each successor n′ of n do
5:    if n′ is not in any of the lists then
6:       Level(n′) ← Level(n) + 1; Insert n′ to ChildrenList;
7:    else if g(n′) < its previous g-value then
8:       Update Level(n′), Parent(n′), g(n′), f(n′); Move n′ to ChildrenList;
9:    if |ChildrenList| > K then
10:      Move a node with largest f-value from ChildrenList to SuspendList;
```

Finally, Algorithm 3 presents the pseudo-code of the `GenerateChildren` routine. It generates the children within the given depth bound MAX_DEPTH. Each child is checked as to whether it is already present in the memory, updated with the shortest path from start node, and it is inserted/moved to *ChildrenList*, similar to what is done by all the search algorithms. Whenever the size of *ChildrenList* exceeds K, a least promising node is moved to *SuspendList*, which is a feature specific to APS.

Additionally, when using admissible heuristics, one can prune the nodes whose f-value exceeds that of the best known solution. Next, we present the properties satisfied by the proposed method.

Property 1: APS is complete and guarantees termination with an optimal solution, provided MAX_DEPTH is at-least as large as the number of nodes on a minimum-length optimal solution path.

Property 2: APS expands at-most $K \times MAX\_DEPTH$ number of nodes in each iteration, where K is the pack size.

The above property holds since the depth of shallowest nodes in *OpenList* increases by 1 in each step of the iteration (to a maximum of $MAX\_DEPTH - 1$), and each step involves expansion of at-most K nodes.

An interesting variation of the proposed algorithm is **Anytime Pack Progressive Search (APPS)**. In this variation, we initialize the value of *pack size* K with $INIT$ and increase it in each iteration by $STEP$ while it is less than $BOUND$. Typical values of $INIT$, $STEP$, and $BOUND$ are 1, 1, and 100 (or some constant) respectively. $BOUND$ helps in converging to the solutions in a time-bounded manner.

Another useful variation is named **Anytime Pack Scaling Search (APSS)**. This variation proceeds in a similar manner to APPS, except that whenever a better solution is found by the algorithm, the *pack size* K is reset to $INIT$, with the hope that the next better solution may be found with only a minimal effort (number of node expansions).

3 Experimental Results

Now, we present the results comparing APS and its variations against Depth-first Branch and Bound (DFBB) [5], Beam-Stack search (BS) [10], Anytime Window A* (AWA*) [1], and Anytime Non-parametric A* (ANA*) [9]. Since the performance of Beam-Stack algorithm varies with the beam-width, we have tested it with multiple values and presented the best results obtained amongst all of them. All the experiments have been performed on a Dell Precision T7600 Tower Workstation with Intel Xeon CPU E5-2687W at 3.1-GHz $\times$ 16 and 256-GB RAM. We display the results in terms of a metric called % *Optimal Closeness* (measures closeness to the optimal solution for minimization problems; whenever optimal solution is known), which is defined as: % *Optimal Closeness* = (*Optimal solution*/*Obtained solution*) $\times$ 100.

Sliding-tile Puzzle Problem (SPP): For our experiments, we have considered all 50 24-puzzle instances from [4, Table II]. Manhattan distance heuristic

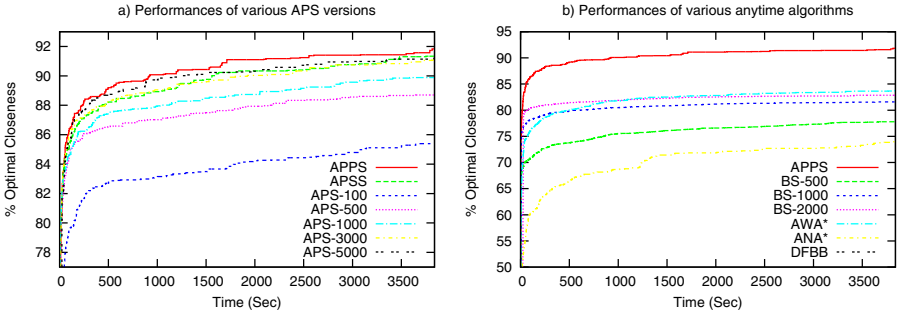


Fig. 1. Comparison of the average of anytime performances on the 50 Puzzle instances. DFBB could not come up with any solution in most cases in the given time.

is used as the heuristic estimation function (which underestimates the actual distance to goal). All algorithms are allowed to explore up-to a maximum depth of 1000 (around 10 times the depth of optimal solutions).

Figure 1a shows the comparison of anytime performances of different APS versions on the Puzzle instances. Here APPS and APSS are run with $INIT$, $STEP$, and $BOUND$ being 1000, 1000, and 5000 respectively. APS-x denotes that APS is run with pack size x. Figure 1b shows the comparison of anytime performance of best performing APS version– APPS, with that of the other anytime algorithms where it can be seen that APPS clearly stands out better. Here, BS-x denotes that beam-stack search is run with beam-width x.

Traveling Salesman Problem (TSP): We chose the first 50 symmetric TSPs (when sorted in increasing order of their sizes) from the traveling salesman problem library (TSPLIB) [8] for our experiments. These range from burma14 to gr202 where the numerical postfixes denote the size of the TSPs. Minimum spanning tree (MST) heuristic is used as the heuristic estimation function (which is an under-estimating heuristic).

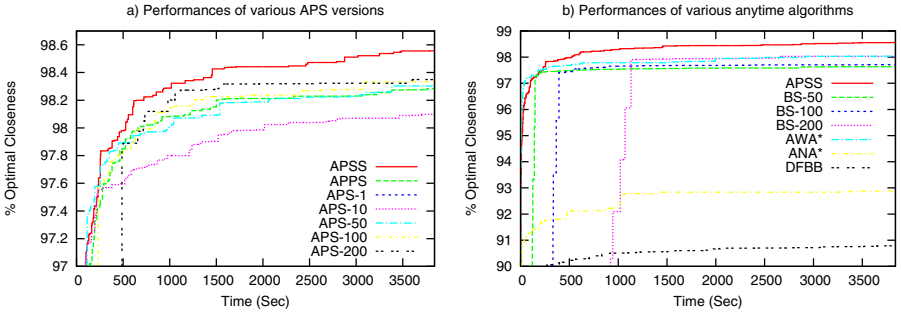


Fig. 2. Comparison of the average of anytime performances on the 50 TSP benchmarks

Figure 2a shows the comparison of anytime performances of different APS versions on the TSP instances. Here APPS and APSS are run with $INIT$, $STEP$, and $BOUND$ being 1, 1, and ∞ (no bound) respectively. Figure 2b shows the comparison of anytime performance of best performing APS version– APSS, with that of the other anytime algorithms where once again it can be seen that the APS version outperforms the others.

Single-Machine Scheduling Problem (SMS): We have generated 48 SMS problem instances[1] with different tardy factors and range factors using similar settings as given in [6]. Each instance contains 100 jobs. For each job j, an integer processing time p_j, earliness penalty h_j, and tardiness penalty w_j are generated from the uniform distribution $[1, 10]$. Let P be the sum of processing times of all the jobs. For each job j, an integer due date is generated from the uniform distribution $[P(1 - T - R/2), P(1 - T + R/2)]$, where T is

[1] The SMS benchmarks used in this paper are available at `https://sites.google.com/site/satyagautamv/resources/sms-benchmarks`

the tardy factor, set at $0.0, 0.2, 0.4, 0.6, 0.8$ and 1.0, and R is the range factor, set at $0.2, 0.4, 0.6$ and 0.8. For each combination of parameters, two problems are generated. The lower bound computation method proposed in [6] is used for computing the under-estimating heuristic over the unscheduled jobs. Here, since the optimal solutions are not available, we display the results in terms of a metric called % *Comparative Closeness* (measures closeness to the best known solution), which is defined as: % *Comparative Closeness* = (*Best Known solution*/*Obtained solution*) $\times$ 100. We obtained the best known solution for each instance by finding the best result produced amongst all the algorithms tested by us for that instance.

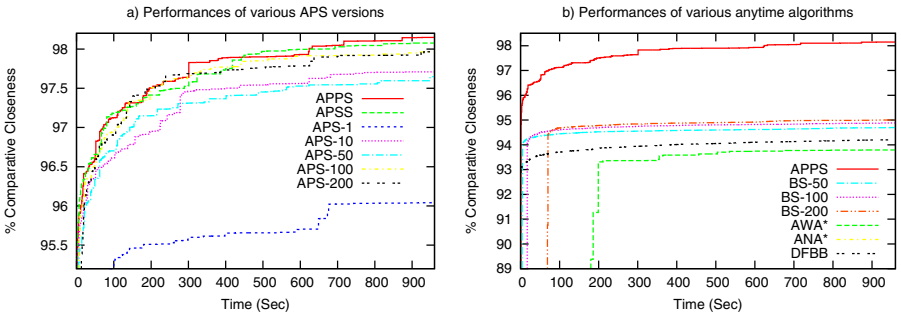


Fig. 3. Comparison of the average of anytime performances on the 48 SMS problem instances. ANA* could not come up with any solution in most cases in the given time.

Figure 3a shows the comparison of anytime performances of different APS versions on the SMS problem instances. APPS and APSS are run with $INIT$, $STEP$, and $BOUND$ being 1, 1, and ∞ (no bound) respectively. Figure 3b shows the comparison of anytime performance of best performing APS version–APPS, with that of the other anytime algorithms where the APS version dominates the others.

In conclusion, the proposed anytime heuristic search technique shows a lot of promise in solving hard combinatorial problems of both AI and Optimization domains. The choice of K (pack size) often provides a trade-off between the time taken for producing initial solution and the quality of anytime performance later on. APPS and APSS take advantage of faster initial solutions (obtained using lower K values) and better later solutions (obtained using higher K values) while avoiding the individual defects of smaller or larger K values and also making the algorithm independent of the choice of K. Experimental results over the three domains considered show that the progressive/scaling algorithms perform better than the other versions (using a specific K) and the existing algorithms.

References

1. Aine, S., Chakrabarti, P.P., Kumar, R.: AWA* - A window constrained anytime heuristic search algorithm. In: Veloso, M.M. (ed.) IJCAI, pp. 2250–2255 (2007)
2. Chakrabarti, P.P., Aine, S.: New approaches to design and control of time limited search algorithms. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) PReMI 2009. LNCS, vol. 5909, pp. 1–6. Springer, Heidelberg (2009)
3. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Transactions on Systems Science and Cybernetics 4(2), 100–107 (1968)
4. Korf, R.E., Felner, A.: Disjoint pattern database heuristics. Artif. Intell. 134(1-2), 9–22 (2002)
5. Lawler, E.L., Wood, D.E.: Branch-and-bound methods: A survey. Operational Research 14(4), 699–719 (1966)
6. Li, G.: Single machine earliness and tardiness scheduling. European Journal of Operational Research 96(3), 546–558 (1997)
7. Lowerre, B.: The Harpy Speech Recognition System. PhD thesis, Carnegie Mellon University (1976)
8. Reinelt, G.: TSPLIB - A traveling salesman problem library. ORSA Journal on Computing 3, 376–384 (1991)
9. van den Berg, J., Shah, R., Huang, A., Goldberg, K.Y.: Anytime nonparametric A*. In: AAAI, pp. 105–111 (2011)
10. Zhou, R., Hansen, E.A.: Beam-stack search: Integrating backtracking with beam search. In: Proceedings of the 15th International Conference on Automated Planning and Scheduling (ICAPS 2005), Monterey, CA, pp. 90–98 (2005)

A Subspace Module Extraction Technique for Gene Expression Data

Priyakshi Mahanta[1], Dhruba Kr. Bhattacharyya[1], and Ashish Ghosh[2]

[1] Department of Comp. Sc. and Engg., Tezpur University, Napaam -784028, India
[2] Machine Intelligent Unit, Indian Statistical Institute, Kolkata-700108, India
{priyakshi,dkb}@tezu.ernet.in,
ash@isical.ac.in

Abstract. Construction of co-expression network and extraction of network modules have been an appealing area of bioinformatics research. In literature, most existing algorithms of gene co-expression network extract network modules where all samples are considered. In this paper, we propose a method to construct a co-expression network based on mutual information and to extract network modules defined over a subset of samples. The method was applied over several real life gene expression datasets and the results are validated in terms of p value, Q value and topological properties.

Keywords: Co-expression network, network modules, mutual information, topological property.

1 Introduction

Mutual information is a measure of similarity between two variables. It has been widely used in various application domains including computational biology, machine learning, statistics, image processing, and financial computing. Mutual information (MI) provides a general measurement for dependencies in the data, in particular positive, negative and nonlinear correlations between two variables. An advantage of MI compared to many other similarity measures such as Pearson correlation and Euclidian distance, is its ability to detect non-linear correlations between two variables.

The concept of mutual information was initially developed for discrete data. The Mutual Information(MI) between two continuous random variable X and Y is defined as follows:

$$I(X;Y) = \int \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} \, \mathrm{d}x \, \mathrm{d}y. \tag{1}$$

where *f(x,y)* is the joint probability density of the two random variables and *f(x)* and *f(y)* are the marginal densities. Computing MI value for all possible pairs of genes is a computationally expensive task. In this paper, we use the

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 635–640, 2013.

method proposed by [1] where the order of calculations is rearranged so that repeating operations are avoided significantly.

Mutual Information between two variables can be computed if we have explicit knowledge of the joint probability distribution and marginal probability distribution. Generally these probabilities are not known. Various methods are used to estimate the probability densities from the observed data. Histogram approach and Kernel Estimation [2] are two popular methods used to estimate probabilities from observed data. Histogram approach is sensitive to the choice of the origin. Kernel estimation gives a better mean square error rate compared to histogram approach. Given M data points drawn from the joint probability distribution, (x_j, y_j) where *j=1, 2, 3,...., M,* the joint and marginal probability can be estimated by gaussian kernel estimator where h is a tuning parameter that determines the width of the kernel.

$$f(x,y) = \frac{1}{M}\sum \frac{1}{2\pi h^2} e^{-\frac{1}{2h^2}(x-x_j)^2+(y-y_j)^2}. \tag{2}$$

$$f(x) = \frac{1}{M}\sum \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{1}{2h^2}(x-x_j)^2}. \tag{3}$$

$$f(y) = \frac{1}{M}\sum \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{1}{2h^2}(y-y_j)^2}. \tag{4}$$

Using Gaussian kernel, mutual information between two random variables X and Y can be computed as

$$f(x,y) = \frac{1}{M}\sum_i \log \frac{M\sum_j e^{-\frac{1}{2h^2}((x-x_j)^2)+(y-y_j)^2)}}{\sum_j e^{(-\frac{1}{2h^2})(x-x_j)^2}\sum_j e^{(-\frac{1}{2h^2})(y-y_j)^2}}. \tag{5}$$

In this work, we have attempted to propose a method MIMET (Mutual Information based Module Extraction technique) to construct co-expression network that is capable of reflecting non-liner correlation among genes over subspaces of samples and to extract network modules, thereof.

2 Method

The method consists of two steps. In the first step, it extract network modules from the co-expression network based on mutual information. In the second step, subspace network modules are extracted using a discretization [3] process.

2.1 Extraction of Network Modules

Co-expression network is constructed by computing kernal based MI of each pair of genes. For each pair of genes, we need an appropriate value of h, a user defined parameter. Computing h for each pair of genes is a time consuming process. Therefore, we propose a way to compute h which works for all pairs of genes. To compute h, we first compute standard deviation of each gene considering all

conditions in the gene expression dataset. Finally, h is computed as standard deviation of the deviations of all genes. While constructing a co-expression network, say G based on D, the proposed technique initially computes the adjacency matrix Adj by using a soft thresholding approach. It encodes edge information for each pair of nodes in the co-expression network from the distance matrix. The distance matrix is obtained by using MI. The maximum spanning tree is extracted from the adjacency matrix using Prims algorithm [4] with weights defined by mutual information. The weakest edge in the spanning tree in terms of weight is removed to break the tree into two parts or subtrees. Each subtree is further processed to extract finer (i.e. strongly connected) module.

2.2 Extraction of Subspace Network Modules

In traditional network modules, the edges are defined over the whole set of samples. The subspace network modules are defined over subset of samples [5]. Each of the initial network module is processed individually to extract the subspace network modules. For each condition c_i, it identifies three subsets of genes: (i) subset with larger expression level, (ii) subset of genes with smaller expression level and (iii) subset of genes with similar expression level for condition c_{i+1} with respect to condition c_i. Next it propagates through rest of conditions to filter the subsets in terms of regulation patterns with respect to selected conditions by including more conditions to finally form the subspace modules. If a subset of genes found in the earlier sub-step is unique with respect to it's expression pattern over the subset of conditions (i.e. not found in previous step), accept this subset of genes as a discovered set of subspace network modules which fulfils the criteria of minimum number of genes in a subspace module.

The complexity of finding MI matrix is O (n^2m^2) where n and m are the number of gens and conditions in a dataset. Finding a maximum spanning tree consumes a complexity of O($n_c g_a^2$) where n_c is the number of network modules and g_a is the average number of genes in the module. The complexity of finding subspace network modules is approximately of O(g_{sub}) where g_{sub} is filtered average number of genes in terms of selected conditions in a module.

3 Experimental Results

We implemented MIMET algorithm in MATLAB and tested it on three benchmark microarray datasets mentioned in Table 1. The test platform was a Sun workstation with Intel(R) Xenon(R) 3.33 GHz processor and 6 GB memory running Windows XP operating system.

3.1 Validation

For biological validation of the extracted network modules, we used well known p and Q value. Also to establish effectiveness of MIMET, we used topological validation.

Table 1. Datasets used for evaluating *MIMET*

Serial. No	Dataset	No. of Genes/ No. of Conditions	Source
1	Yeast Sporulation	474/17	http://cmgm.standford.edu/pbrown/sporulation
2	Human Dataset	5008/4	Sample gene in expander
3	Rat CNS	112/9	http://faculty. washington.edu/kayee/cluster

(a) p value: Biological significance of the sets of genes included in the extracted network modules are evaluated based on p values [6] which signifies how well these genes match with different Gene Ontology(GO) categories. To compute p-value, we used a tool called FuncAssociate [7]. (b) Q value: Q-value [8] for

Table 2. p-values of a network module extracted for Dataset 1

P-value	*GO number*	*GO category*
3.38e-31	GO:0048646	anatomical structure formation involved in morphogenesis
5.89e-27	GO:0042244	spore wall assembly
3.13e-23	GO:0048869	cellular developmental process
3.81e-20	GO:0070882	cellular cell wall organization or biogenesis
3.81e-20	GO:0071554	cell wall organization or biogenesis
1.06e-25	GO:0043934	sporulation
6.10e-29	GO:0010927	cellular component assembly involved in morphogenesis
1.06e-25	GO:0030154	cell differentiation
3.53e-28	GO:0070726	cell wall assembly
5.89e-27	GO:0030476	ascospore wall assembly
5.72e-16	GO:0005628	prospore membrane
5.89e-27	GO:0071940	fungal-type cell wall assembly
2.06e-22	GO:0032502	developmental process
2.84e-17	GO:0048610	cellular process involved in reproduction
1.06e-15	GO:0006094	gluconeogenesis
6.60e-19	GO:0070882	cellular cell wall organization or biogenesis
1.43e-25	GO:0030435	sporulation resulting in formation of a cellular spore

a particular gene G is the proportion of false positives among all genes which are more differentially expressed. We used a web based tool GeneMANIA [9] to generate hypotheses about gene function, analyzing gene lists and prioritizing genes for functional assays. From the given p and Q values shown in Tables 2 and 3, we can conclude that MIMET shows a good enrichment of functional categories and therefore project a good biological significance.

Topological Validation. The meaning of the edges in a gene co-expression network is a relevant question in network analysis. Different structural properties of co-expression network can be a solution to the above question. Therefore, in this paper the structural properties of the co-expression network inferred from gene expression microarray data are compared with topological properties of the known, well established network data of the same organism. We use a web based tool called topoGSA [10] for topological validation of the extracted network modules. MIMET has been found capable of identifying gene modules with very high average score **($\geq$.85)** as shown in Table 4.

Table 3. Q-value of one of the network modules of Dataset 3

GO annotation	*Q value*
regulation of transmission of nerve impulse	2.50E-11
axonogenesis	2.50E-11
regulation of synaptic transmission	2.50E-11
regulation of neuron apoptosis	2.50E-11
axon	2.50E-11
regulation of neurological system process	4.20E-11
neuron apoptosis	4.20E-11
neuron death	5.33E-11
negative regulation of neuron apoptosis	1.56E-10
transmembrane receptor protein tyrosine kinase signaling pathway	2.13E-10
growth factor binding	4.40E-9
protein autophosphorylation	1.73E-8

Table 4. Ranking of KEGG gene sets based on topological similarity to uploaded module for dataset 2

Identifier	*Median* degree	*Median* CC	*Median* SPL	*Median* BW	*Median* EVC	*Score*
hsa05216:Thyroid cancer	20	0.07	3.45	26256	0.05	0.83
hsa05210:Colorectal cancer	22.5	0.07	3.44	31963	0.05	0.84
hsa04520:Adherens junction	26	0.05	3.38	33514	0.07	0.85
hsa04012:ErbB signaling pathway	24	0.07	3.39	28589	0.07	0.87
hsa05214:Glioma	28	0.05	3.35	37425	0.07	0.87
hsa05213:Endometrial cancer	24	0.07	3.38	33066	0.07	0.87
hsa01510:Neurodegenerative Diseases	33	0.04	3.33	67612	0.05	0.88
hsa05215:Prostate cancer	28	0.06	3.37	35660	0.07	0.89
hsa05221:Acute myeloid leukemia	27	0.09	3.39	32521	0.07	0.9
hsa04320:Dorso-ventral axis formation	30	0.06	3.32	32814	0.1	0.91
hsa05223:Non-small cell lung cancer	29	0.07	3.35	38253	0.08	0.91
hsa05212:Pancreatic cancer	30	0.07	3.35	45788	0.07	0.92
hsa05220:Chronic myeloid leukemia	36	0.08	3.3	49218	0.1	0.96

4 Conclusion and Future Work

In this paper, a gene co-expression network construction technique called MIMET based on a soft thresholding approach is presented. The performance of MIMET is excellent in three publicly available benchmark real life datasets. From the network, subspace network modules are extracted using a discretization process. An adaptive discretization technique can also be used to extract more biologically effective network modules. Construction of gene regulatory network from co-expression network with the support of Gene Ontology can lead to better, more meaningful predictions of gene networks.

Acknowledgment. This paper is an outcome of a research project supported by DST, Govt. of India in collaboration with CSCR, ISI, Kolkata.

References

[1] Qiu, P., Gentles, A.J., Plevritis, S.K.: Fast calculation of pairwise mutual information for gene regulatory network reconstruction. Comput. Methods Prog. Biomed. 94(2), 177–180 (2009)

[2] Moon, Y.I., Rajagopalan, B., Lall, U.: Estimation of mutual information using kernel density estimators. Phys. Rev. E 52(3) (September 1995)
[3] Mahanta, P., Ahmed, H.A., Kalita, J.K., Bhattacharyya, D.K.: Discretization in gene expression data analysis: a selected survey. In: Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology, CCSEIT 2012, pp. 69–75. ACM, New York (2012)
[4] Prim, R.C.: Shortest connection networks and some generalizations. Bell System Technology Journal 36, 1389–1401 (1957)
[5] Sharma, S., Bhattacharyya, D.K.: An effective technique for clustering incremental gene expression data. International Journal of Computer Science Issues 7 (2010)
[6] Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M.: Systematic determination of genetic network architecture. Nature Genetics (1999)
[7] Berriz, G.F., King, O.D., Bryant, B., Sander, C., Roth, F.P.: Characterizing gene sets with FuncAssociate. Bioinformatics (Oxford, England) 19, 2502–2504 (2003)
[8] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300 (1995)
[9] Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G.D., Morris, Q.: The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Research 38, W214–W220 (2010)
[10] Glaab, E., Baudot, A., Krasnogor, N., Valencia, A.: TopoGSA: network topological gene set analysis. Bioinformatics 26(9), 1271–1272 (2010)

Simultaneous Gene Selection and Cancer Classification Using a Hybrid Intelligent Water Drop Approach

Manish Kumar[1], Shameek Ghosh[2], Jayaraman Valadi[2,3,*], and Patrick Siarry[4,*]

[1] Bioinformatics Center, University of Pune, Pune, India
rishimanish123@gmail.com
[2] Centre for Development of Advanced Computing, Pune, India
shameekg@cdac.in
[3] Shiv Nadar University, Uttar Pradesh – 203207, India
jayaraman.valadi@snu.edu.in
[4] Université Paris-EstCréteil, Val-de-Marne, LiSSi (EA 3956), France
siarry@u-pec.fr

Abstract. Computational Analysis of gene expression data is extremely difficult, due to the existence of a huge number of genes and less number of samples (limited number of patients). Thus,it is of significant importance to provide a subset of the most informative genesto a learning algorithm, for constructing robust prediction models. In this study, we propose a hybrid Intelligent Water Drop (IWD) - Support Vector Machines (SVM) algorithm, with weighted gene ranking as a heuristic, for simultaneous gene subset selection and cancer prediction. Our results, evaluated on three cancer datasets, demonstrate that the genes selected by the IWD technique yield classification accuracies comparable to previously reported algorithms.

Keywords: Gene Selection, Cancer Classification, Intelligent Water Drop based Optimization, Weighted Ranking.

1 Introduction

The number of genes in a microarray gene expression dataset is normally much greater than the number of samples (instances), which makes the disease prediction problem difficult to solve since, out of thousands of genes, most genes do not correlate with the prediction process. To improve model accuracy, it is thus important to select a subset of relevant genes from the data. This is known as *Gene Selection or Feature Selection* and it helps in getting rid of irrelevant and noisy genes [1]. Two important categories of gene selection methods are: 1) *wrappers* and 2) *filters* [1]. Wrappers use a learning algorithm to score the quality of gene subsets based on their predictive power. On the other hand, filters select subsets of genes independently of the chosen predictor and evaluate the quality of genes considering their statistical properties. The problem of gene selection typically falls into the category of large-scale global

* Corresponding authors.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 641–646, 2013.

optimization. Various nature-inspired optimization algorithms have been reported to solve such problems. Among these, swarm intelligence based methods have their own niche and sophisticated operators, which explore novel swarm based approaches to tackle optimization problems. According to the No-Free-Lunch Theorem [11], all metaheuristic based approaches report the same performance results when averaged over all possible objective functions. Thus, even though the spectrum of gene selection problems is quite huge, the numbers of reported swarm based metaheuristics are much less in comparison. Hence, in this study, we present a hybrid Intelligent Water Drop Optimization (IWD) based filter-wrapper approach for selecting a relevant subset of genes most predictive of a certain type of cancer.

2 Methodology

2.1 Intelligent Water Drop Based Optimization

The Intelligent Water Drop (IWD) algorithm has been inspired by the study of the real behavior of natural drops in a flowing water source from high altitude to low altitude regions. Shah-Hosseini extended this natural concept to introduce the Intelligent Water Drop (IWD) algorithm for the Travelling Salesman Problem (TSP) [2]. An IWD consists of two major properties - 1) the soil content of the IWD - *soil(IWD)* and 2) the velocity of the IWD - *vel(IWD)*. The IWD soil and velocity content dynamically change based on the path taken by the same, while flowing through the discrete problem landscape. Depending on the IWD movement, some soil is thus removed from the traversed path and the corresponding path soil is updated dynamically in the process. Such a flow results in the lowering of soil content in optimal routes based on the problem environment. One can thus say that the paths with lesser soil content may be the most relevant for the search of a near optimal solution. Based on the original formulation for a TSP problem, we may consider a graph G= (V, E) where V is the set of nodes and E, the set of edges. An IWD can thus be randomly placed at any node (say i). To select the next node (j), it follows the probability transition as given in equation (1).

$$P(i,j) = \frac{f(soil(i,j))}{\sum_{kisunvisited} f(soil(i,k))} \tag{1}$$

$$f(soil(i,j)) = \frac{1}{\varepsilon + g(soil(i,j))} \tag{2}$$

$$g(soil(i,j)) = \begin{array}{l} soil(i,j) \text{ if } minsoil >= 0 \\ soil(i,j) - minsoil \text{ if } minsoil < 0 \end{array} \tag{3}$$

Here P(i,j) indicates the transition probability associated with node j. k specifically denotes all the nodes that are still to be visited. ε is an algorithmic parameter. Thus the selection of a node depends probabilistically on the amount of soil present on the edges between adjacent nodes given by soil(i,j). Here *minsoil* indicates the least soil available on a path between any node i and j. As illustrated in equations (1) to (3), the state transition probability of an IWD is thus proportional to the soil content available in the edge between nodes i and j. While each IWD incrementally moves from one

node i to j while constructing a solution, the IWD soil content (soil(iwd)) and the velocity of the same(vel(iwd)) are also updated based on equations (4-5).

$$\Delta vel^{IWD}(t) = \frac{a_v}{b_v + c_v \times soil^{2\alpha}(i,j)} \tag{4}$$

$$\Delta soil(i,j) = \frac{a_s}{b_s + c_s \times time^{2\theta}(i,j)} \tag{5}$$

Here the IWD velocity is changed by a Δ vel component. a_v ,b_v ,c_v , and α are algorithm specific parameters. Similarly the soil content of an IWD is also increased by Δsoil which is the soil content removed by the IWD while moving from location i to j. $time^{2\theta}(i,j)$ is the time required for the IWD to move from i to j which is given as –

$$time(i,j) = \frac{HUD(i,j)}{vel(IWD)} \tag{6}$$

HUD is characterized as a heuristic which can be used to measure the desirability/undesirability of an IWD to select an edge between i and j. θ, in this case, is an algorithmic parameter. Thus, a larger IWD velocity contributes to minimizing the time taken by an IWD to move from i to j. The time factor in turn influences the amount of soil to be removed from a path (as shown in equation 5). Once the IWD properties are computed, the soil content of the complete solution path can be updated based on equation 7.

$$soil(i,j) = \rho_0 \times soil(i,j) - \rho_n \times \Delta soil(i,j) \tag{7}$$

whereρ_o and ρ_n are between 0 and 1.According to the original IWD algorithm for the TSP, $\rho_o = 1 - \rho_n$.

2.2 IWD Based Feature Selection

For the feature selection problem, we consider each node (in the graph above) as a feature. Thus if a gene expression dataset consists of 1000 features, then a possible solution could be a feature subset composed of {11,23,391,510,999} with the subset size as 5. Here each element is a feature index. An initial set of IWDs are thus placed at random features from where they commence their flow. Each IWD moves to the next feature by following the probability transition given by equation (1). Once a feature has been visited, a local soil update between features i and j are performed by equation (7) as mentioned before. In the process, the IWD soil content and velocity are also updated by equation 4-5. This process, continues until a complete feature subset of the required size is constructed by the IWD. The feature subset is then used to generate a corresponding reduced dataset with the given features indices. The reduced dataset is thus fed as input to a classifier like SVM, which consequently returns a 10 fold classification cross validation accuracy (10 fold CVA).The 10 fold CVA is thus considered as the fitness measure for the corresponding feature subset(or the IWD solution). Subsequent IWDs also build up their solution vectors (feature subsets) similarly. After each iteration, the feature subset with the maximum 10 fold CVA gets selected as the iteration best solution (TIB). A certain amount of soil is removed from the edges of the

iteration-best solution based on the quality of the feature subset. Thus if TIB is given as (6,13,91,121,992), then the edges to be updated are 6-13,13-91,91-121 and 121-992.This is done according to equation (8).

$$soil(i,j) = \rho_s \times soil(i,j) - \rho^{IWD} \times \frac{1}{N_{IB} - 1} \times soil_{IB^{IWD}} \tag{8}$$

Here $soil_{IB^{IWD}}$ represents the soil content of the iteration-best IWD (which owns the iteration best feature subset). N_{IB} is the number of features in TIB. ρ^{IWD} is the global soil updating parameter selected from [0, 1]. ρ_s is set as $(1+ \rho^{IWD})$. Therefore an edge with lesser soil content turns out to have better prospects in the future in the constructing a good solution.

In addition we also maintain a global best feature subset which is given by the maximum of all the iteration best solutions. The above process is repeated till a termination criterion is reached. During this stage, the global best feature subset is reported as the most optimal solution to the feature selection problem. The IWD gene selection algorithm is thus stated as below.

2.3 Weighted Gene Ranking

A weighted gene ranking composed of three filters namely Information Gain (IG), Chi-square(CS) and Correlation based feature selection (CFS), are provided as input to the IWD algorithm [3].The heuristic information for each individual gene is obtained by calculating the weighted sum of the IG, CS and CFS scores which were obtained using the WEKA[4] data mining library. The computation of the weighted sum of a gene (WRg) is as shown in equation (9).

$$WR_g = w_1 * IG_g + w_2 * CS_g + w_3 * CFS_g \tag{9}$$

Here, w1,w2 and w3 are the weights provided for IG, CS and CFS rankings. The WRg is consequently provided as HUD(i,j) for the j-th feature, as shown in equation (6), in a modified form as given in equation (10).

$$time(i,j) = \frac{1}{WR_j + vel(IWD)} \tag{10}$$

The weighted gene value (WRj) is thus used to probabilistically guide the IWD search.

2.4 Support Vector Machine

Support Vector Machines (SVM) were introduced by Vapnik et al [5-6] and successively extended by a number of other researchers. SVM uses a maximum margin linear hyperplane for solving binary linear classification problems. For problems that are non-linearly separable, SVM transforms the data into higher dimensional features and then employs a linear hyperplane. To deal with intractability issues it also employs appropriate kernel functions allowing computations in the input space itself. In particular, SVM with recursive feature elimination (RFE) was used by Vapnik et al [7] for gene selection and achieved notably high accuracy levels. For our purposes, we

employ the libSVM [8] library for evaluation of our candidate solutions during each generation.

3 Results and Discussion

Microarray gene expression datasets specify the expression levels of different genes, which are available publicly. Three such datasets were obtained from the Kent Ridge Biomedical datasets repository [8] and the libSVM repository [7].The Colon cancer dataset consists of 2000 genes and 62 samples. The Breast Cancer data constitues 7129 genes and 44 samples. The leukemia dataset consists of 7129 genes and 72 samples. Each of the three datasets constitutes a binary classification problem. Extensive simulations were carried out for each dataset with separate gene rankings as Information gain, Chi-square, CFS and the weighted heuristics as described earlier. Based on the simulations, one can say that comparable results for all three datasets were observed, while considering a maximum of 50 IWDs and 100 generations. Mostly towards the end of 100 generations, the fitness values of the feature subsets would converge and not show much improvement. Parameter tuning was also carried out extensively for weighted ranking to get the best results. Our simulations indicate that the SVM kernel and filter weighting parameters have a more profound influence and have thus tuned the same extensively for maximizing algorithm performance. The algorithm parameters for IWD are as shown in Table II.A comparison of the weighted IWD-SVM performance is provided along with the some recently reported best results for the same datasets. The results of the simulations are as given in Table III.

Table 1. IWD Parameters

IWD Algorithm Parameters	*Values*
No. of IWDs	50
w_1,w_2,w_3	0.5,0.3,0.2
No. of Generations	100
a_v,b_v,c_v,α	1,0.01,1,1
a_s,b_s,c_s,θ	1,1,0.01,2
ρ_o , ε	0.1,0.5
cost,gamma(for radial basis function as SVM kernel),Folds	50,0.02,10

Table 2. Comparison of IWD-SVM with previously reported classification accuracies [1, 9, 10]. ACO-AM: Ant Colony Optimization with AntMiner, ACO-RF: Ant Colony Optimization with Random Forests.

Colon	85.48% (SVM)	95.47%(ACO-AM)[1]	96.77%(ACO-RF) [1]	95.16%(IWD-SVM)
Breast	60.02% (SVM)	92.00% (Bagging)[9]	94.00%(Ensemble Predictors)[10]	97.72 %(IWD-SVM)
Leukmia	94.73% (SVM)	96.00%(ACO-AM)[1]	69.00% (nearest neighbor search)	97.22%(IWD-SVM)

According to results in Table III, IWD-SVM performs well in comparison to previously reported algorithms namely for all the three datasets [1,9,10]. The IWD based gene subset sizes selected were 15 for Colon, 15 for Breast and 19 for Leukemia. In addition, simulations with simple filters like Information Gain, Chi-square ranking and CFS were carried out separately with IWD for similar subset sizes. As per our results, the IWD-SVM with weighted ranking demonstrated superior performances than IWD-Infogain, IWD-Chi-Square and IWD-CFS.

4 Conclusion

The hybrid IWD-SVM has shown good results consistently on comparison with the highest accuracies for colon cancer, breast cancer and leukemia cancer datasets. In general, IWD is robust and flexible for discrete optimization owing to their typical swarm based emergent behavior.

Acknowledgment. VKJ gratefully acknowledges the Council of Scientific & Industrial Research (CSIR), New Delhi, India for financial support in the form of an Emeritus Scientist grant.

References

1. Sharma, S., Ghosh, S., Anantharaman, N., Jayaraman, V.K.: Simultaneous informative gene extraction and cancer classification using ACO-antMiner and ACO-random forests. In: Satapathy, S.C., Avadhani, P.S., Abraham, A. (eds.) Proceedings of the InConINDIA 2012. AISC, vol. 132, pp. 755–761. Springer, Heidelberg (2012)
2. Shah-Hosseini, H.: Problem solving by intelligent water drops. In: IEEE Congress on Evolutionary Computation, CEC 2007, pp. 3226–3231 (2007)
3. Han, J., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann (2006)
4. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
5. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT 1992, pp. 144–152. ACM, New York (1992)
6. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422 (2002)
7. Chang, C.-C., Lin, C.-J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27 (2011)
8. Kent ridge bio-medical dataset, `http://datam.i2r.astar.edu.sg/datasets/krbd/`
9. Martn-Merino, M., Blanco, A., De Las Rivas, J.: Combining dissimilarity based classifiers for cancer prediction using gene expression profiles. BMC Bioinformatics 8 (2008)
10. Cong, G., Tan, K.-L., Tung, A.K.H., Xu, X.: Mining top-k covering rule groups for gene expression data. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2005, pp. 670–681. ACM, New York (2005)
11. Wolpert, D.H., Macready, W.G.: No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1(1), 67–82 (1997)

Scored Protein-Protein Interaction to Predict Subcellular Localizations for Yeast Using Diffusion Kernel

Ananda Mohan Mondal[1,*] and Jianjun Hu[2]

[1] Mathematics and Computer Science, Claflin University, Orangeburg, USA
amondal@claflin.edu
[2] Computer Science and Engineering, University of South Carolina, Columbia, USA
jianjunh@cec.sc.edu

Abstract. Network-based protein localization prediction is explored utilizing the protein-protein interaction score along with the network connectivity. Score-based diffusion kernel is introduced to solve the problem. Four different PPI networks, namely, co-expressed PPI, Genetic PPI, Physical PPI, and scored PPI are used for analysis. Our investigation shows that PPI score does have positive impact in predicting subcellular protein localization. At high average PPI score of 891, performance accuracy ranges from 0.78 for 'punctate composite' to 0.93 for 'nucleolus' and at low average PPI score of 169, performance accuracy ranges from 0.60 for 'cytoplasm' to 0.83 for 'mitochondrion'.

Keywords: Scored PPI, subcellular protein localization, protein localization, diffusion kernel, NetLoc.

1 Introduction

Precise targeting to designated subcellular localization is essential for proper protein function. Experimental determination of protein localization is costly [1, 2]. Computational algorithm can greatly help in predicting protein localizations, which in turn can infer protein functions. In the past decade, many computational algorithms have been developed for predicting subcellular localization of protein. These algorithms employ a variety of supervised machine learning techniques including support vector machines[3], neural networks [4], nearest neighbor classifier, Markov models, Bayesian networks [5, 6] etc. The existing prediction algorithms can be divided into four major categories [7] in terms of the evidences used: 1) algorithms based on targeting signals; 2) algorithms considering the preference or bias in terms of amino acid; 3) algorithms using localization information from other annotated proteins with indirect relationships such as functional annotation, phylogenetic profiling, homology and protein-protein interaction; and 4) algorithms that integrate multiple sources of information.

Recently, protein-protein correlation (PPC) networks such as PPI networks [8] and metabolic networks [9] have been used for localization prediction. One of the limitations of these methods that these are not capable of utilizing the inherent network

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 647–655, 2013.

information that naturally appears among proteins [7, 10]. In their recent work, Mondal and Hu [7, 10-12] exploit the PPI network information in predicting subcellular protein localization using diffusion kernel. But these studies do not utilize the PPI score meaning they used score of unity for each PPI. PPI score, ranging between 150 and 999, in the STRING database reflects the confidence of functional association of two different proteins [13]. The higher is the score of a PPI, the higher is the probability of two proteins to be associated with the same function. If two proteins are associated with the same function, they are more likely to be localized at the same subcellular location. So, PPI scores in STRING database can better be utilized in predicting subcellular localization of proteins. It is thus interesting to explore score-based diffusion kernel algorithm in predicting protein subcellular localizations.

2 Diffusion Kernel-Based Logistic Regression using Scored PPI

2.1 Score-Based Diffusion Kernel

Score-based diffusion kernel is derived from weighted adjacency matrix of scored PPI network, where weight on an edge is the score of PPI connecting two proteins. Diffusion kernel *K,* to represent the scored interaction network, is defined using the following equation.

$$K = e^{\{W\}} \tag{1}$$

Where

$$W(i,j) = \begin{cases} w_{i,j} & if\ protein\ i\ interacts\ with\ protein\ j \\ -d_i & if\ prtoein\ i\ is\ the\ same\ as\ protein\ j \\ 0 & otherwise \end{cases}$$

Where d_i is the sum of weights of interactions with protein *i* and $e^{\{W\}}$ represents the matrix exponential of the matrix W. It is noticeable that $w_{i,j} = 1$ represents the diffusion kernel for non-score based PPI network or PPI with score of unity, which is used in [7, 10-12]. Kernel function $K(i, j)$ represents the similarity distance between protein *i* and protein *j* in the network.

2.2 Kernel-Based Logistic Regression (KLR) Model

For classification, we applied the diffusion kernel-based logistic regression (KLR) model [14] as used in [7, 10-12] to predict protein subcellular localization. Given a protein-protein interaction network with N proteins $X_1, \ldots, X_N$ with n of them $X_1, \ldots, X_n$ with unknown subcellular locations, the objective is to find subcellular locations of n unknown proteins using the locations of known proteins and protein-protein interaction network.

Let

$$X_{[-i]} = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_N)$$

$$M_0(i) = \sum_{j \neq i, x_j \text{ known}} K(i,j) I\{x_j = 0\}$$

,

And $M_1(i) = \sum_{j \neq i, x_j \text{ known}} K(i,j) I\{x_j = 1\}$,

where $K(i,j)$ is the kernel function derived in section 2.2. Indicator $I(x_j = 0)$ represents that the interacting protein *j* does not have the location of interest and indicator $I(x_j = 1)$ represents that protein *j* does have the location of interest. Upon simplification, the KLR model is given by:

$$\log \frac{\Pr(X_i = 1 \mid X_{[-i]}, \theta)}{1 - \Pr(X_i = 1 \mid X_{[-i]}, \theta)} = \gamma + \delta M_0(i) + \eta M_1(i)$$

This means that the logit of $(X_i = 1 | X_{[-i]}, \theta)$, the probability of a protein targeting a location is linear based on the summed distances of proteins targeting to that location or other locations. This equation can be rewritten as

$$\Pr(X_i = 1 \mid X_{[-i]}, \theta) = \frac{1}{1 + e^{-(\gamma + \delta M_0(i) + \eta M_1(i))}}$$

Maximum likelihood estimation (MLE) method is used for estimating the parameters $\gamma, \delta,$ and η. It is noticeable that only the annotated proteins are used in the estimation procedure.

Fig. 1 presents the flow diagram of the scored PPI network-based framework for protein localization prediction employing the KLR model. First, a scored PPI network or weighted adjacency matrix is obtained from the list of scored PPIs. Then diffusion kernel is determined using eqn. (1).

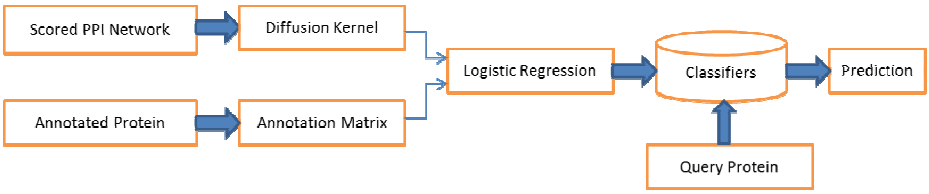


Fig. 1. Protein localization prediction using scored PPI network employing kernel-based logistic regression

Annotation matrix is developed from annotated proteins. This is an *m* by *n* matrix, consists of 1 (annotated) and 0 (not annotated), where *m* is the number of annotated proteins and *n* is the number of localizations. Finally, KLR model is developed from diffusion kernel and annotation matrix using logistic regression.

2.3 Performance Evaluation

The outputs of the KLR model are confidences for each protein to be localized at each of the locations. A threshold on confidence value is used to classify the proteins to be localized at a location or not. If the threshold is set to 0.7, then a protein with higher than 0.7 confidence will be labeled as positive prediction meaning that the protein belongs to this location, otherwise, negative prediction. The results of the KLR prediction algorithm can have varying true positive and true negative rate depending on the threshold value. This makes the comparison difficult. To avoid this difficulty, the AUC (Area Under the Curve) score was used to measure the prediction capability of the proposed KLR model. 5-fold cross-validation was used to calculate the AUC values.

3 Datasets

3.1 PPI Networks

Four different PPI networks for yeast are used in the present study: two networks, physical PPI and genetic PPI, are obtained from BioGRID database [15], one Scored PPI network from STRING database [13], and one co-expressed PPI network is derived from gene expression data of Stanford University [16]. Pearson correlation is used to derive co-expressed PPI from gene expression data. In this study, the networks are named as co-expressed PPI as COEXP, genetic PPI as GPPI, physical PPI as PPPI, and scored PPI as SPPI. Table-1 summarizes the topology of four network datasets. SPPI is the largest network (proteins: 6314; edges: 489934) and COEXP is the smallest (proteins: 2004; edges: 11954). SPPI is also the densest network (77.6 PPI/protein) followed by GPPI (19.73 PPI/protein), PPPI (9.31 PPI/protein), and COEXP (5.96 PPI/protein).

Table 1. Topology of protein-protein interaction networks

Property	**COEXP**	**GPPI**	**PPPI**	**SPPI**
No. of proteins	2004	5252	5477	6314
Edges	11954	103631	50997	489934
Average interactions per node	5.96	19.73	9.31	77.60

3.2 Annotated Proteins

The developed KLR model is applied to protein localization prediction of yeast proteins using the localization data of Huh et al. [1] as the basis for annotation. After removing ambiguous localization, we have 3919 proteins with 5191 localizations. Out of 22 locations, only 7 locations have more than 100 proteins with known subcellular

localization annotation. These locations are cell periphery, cytoplasm, ER (endoplasmic reticulum), mitochondrion, nucleolus, nucleus, and punctuate composite. We evaluated our network prediction model based on these 7 locations.

3.3 Networks in Terms of PPI Score

In COEXP, GPPI, and PPPI network all PPIs have the equal score of unity even though two proteins of different PPIs have different level of confidence of functional association. In order to investigate the relationship of these three types of PPIs with respect to functional association, we identified the PPIs for these three networks that are common with the STRING PPIs. Fig. 1 presents the cumulative distribution of PPI with PPI score for three different types of PPI. In case of GPPI, 50% PPIs have score larger than 375; in case of PPPI, 50% PPIs have score larger than 800; in case of COEXP 50 % PPIs have score larger than 850. So, two proteins in PPI of PPPI and COEXP network are more likely to be related to the same function. By definition, physical PPI means two proteins physically interact together to produce some products, and co-expressed PPI means two proteins have same level of expression in the same direction and as such they are more likely to be associated with the same function, which results in high PPI score for PPPI and COEXP.

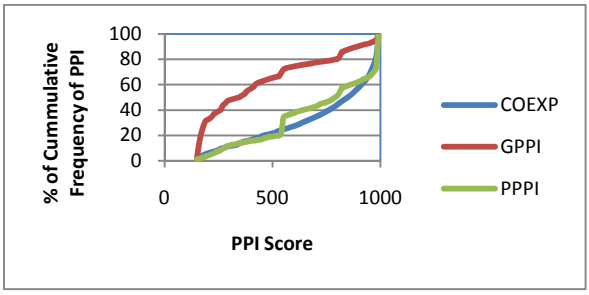


Fig. 2. Distribution of PPIs with PPI score for different types of PPI

4 Results and Discussion

In their previous work [7, 10-12] Mondal and Hu used NetLoc model to predict subcellular localization using PPI network without score. In the present work, we used NetLoc to explore it's capability of similar prediction but using scored PPI from STRING database [13].

In order to see the effect of PPI score in predicting subcellular localization, NetLoc model was applied to five equally divided scored PPI (SPPI) network. The PPIs in the whole SPPI network was sorted in descending order based on PPI score and then divided into five equal parts in terms of number of PPIs. Table 2 summarizes the topology of the divided networks along with the average PPI score and the number of annotated proteins corresponding to each of the divided network. It is noticeable that

topology of the divided networks are very similar in terms of every component of topology: number of PPI ranging from 97986 to 97990, number of protein ranging from 5394 to 5948, and average degree ranging from 33 to 36. The number of corresponding annotated protein is also similar ranging from 3728 to 3870. The major different in the five networks are average PPI score ranging from 169 to 891.

Table 2. Topology of 5 divided networks and average PPI score and number of annotated proteins for corresponding networks

Parts of Network	Network Topology			Avg Score	Ann Protein
	PPI	Protein	Avg Degree		
N1_5	97986	5394	36	891	3728
N2_5	97986	5948	33	575	3870
N3_5	97986	5860	33	359	3837
N4_5	97986	5944	33	230	3839
N5_5	97990	5809	34	169	3809

4.1 Model Performance in Predicting Individual Subcellular Localization

Fig. 3 presents the model performance in predicting 7 subcellular localizations for two of five equally divided networks, N1_5 with the largest average PPI score of 891 and N5_5 with the smallest average PPI score of 169. At high score, performance accuracy ranges from 0.78 for 'punctate composite' to 0.93 for 'nucleolus' and at low score, performance accuracy ranges from 0.60 for 'cytoplasm' to 0.83 for 'mitochondrion'. It is clear that for all locations, network N1_5 performs better than network N5_5. This demonstrates that PPI score does have influence in predicting subcellular localizations. It can be concluded that higher is the average PPI score better is the NetLoc performance.

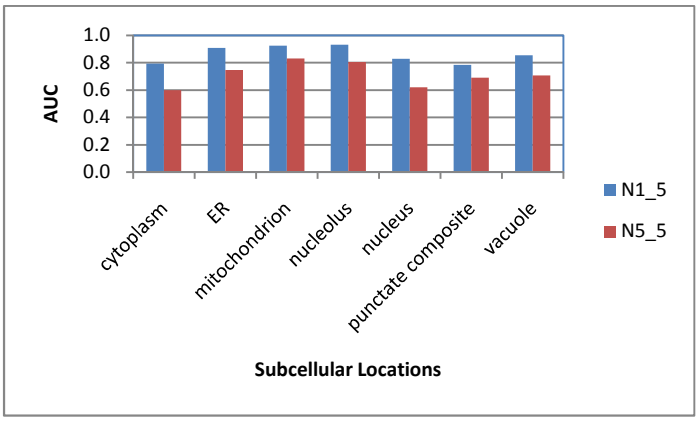


Fig. 3. NetLoc performance in predicting individual subcellular localization

4.2 Overall Performance with Scored PPI Networks

Fig. 4 shows the overall performance for five equally divided networks. It is clear that higher is the PPI score of the network better is the overall performance. This experiment definitely demonstrates that PPI scores have direct impact in predicting subcellular localization. PPI score in the STRING database [13] represents the level of confidence that two proteins can have the similar function. Higher is the score higher is the probability that two proteins are more likely to have the similar function. If two proteins have similar function they are more likely to be localized at the same subcellular compartment [7]. As a result network with high average PPI score produces better results in predicting subcellular localization.

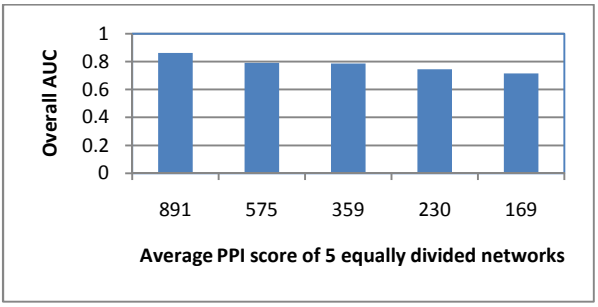


Fig. 3. NetLoc Performance for networks with different PPI scores. Overall AUC is evaluated based on 7 locations

4.3 Comparing Performance with Non-scored PPI Networks

In their previous work [7, 10-12], Mondal and Hu explored protein localization prediction using non-scored PPI networks, namely, COEXP, GPPI, and PPPI. In order to compare we need two networks composed of same number of proteins and same PPIs, one with PPI score and the other without PPI score. As mentioned in section 3.3, these networks can be obtained from the PPIs which are common with STRING PPIs. The derived COEXP contains 8017 PPIs, GPPI contains 85770 PPIs and PPPI contains 45739 PPIs. Fig. 5 shows the model performance for three different types of PPI with and without PPI score. It is evident that improvement upon using PPI score is very small ranging from 1.0% for PPPI to 2.6% for GPPI. This improvement is due to the score attached to each PPI. This proves that network connectivity (prediction without score) provides the most information in predicting protein subcellular localization and inclusion of score features on top of network connectivity have little influence.

4.4 Statistical Significance of Improvement in Prediction due to PPI Score

It is clear from Figs. 3 and 4 that PPI score does have positive impact in predicting protein localization. According to Fig. 5, the improvement in prediction due to score in PPI is very small (1% ~ 3%). Now the question is whether this improvement is statistically significant or it's happening by chance. In Fig. 5, improvement is happening to all three types of PPI. So, it's not happening by chance.

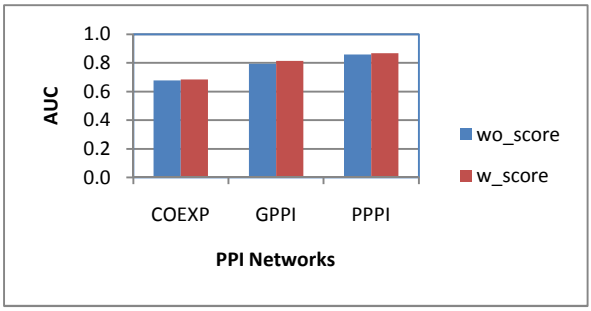


Fig. 4. Overall performance of different types of PPI with and without PPI score

5 Conclusion

A score-based diffusion kernel is introduced in predicting protein subcellular localization using scored protein-protein interaction network. Our investigation shows that PPI scores have direct impact in predicting subcellular localization: higher is the PPI score of the network better is the performance. Our results also show that network connectivity or network with non-scored PPI provides the most information in predicting protein localization. Inclusion of PPI score features on top of network connectivity has little influence.

Acknowledgment. This work was partially supported by NASA grant, Prime Award No: NNX12AI12A, Sub-award No: 520976-Claflin-Mondal and NSF Career Award DBI-0845381.

References

1. Huh, W.K., et al.: Global analysis of protein localization in budding yeast. Nature 425(6959), 686–691 (2003)
2. Agarwal, A.K., et al.: Genome-wide expression profiling of the response to polyene, pyrimidine, azole, and echinocandin antifungal agents in Saccharomyces cerevisiae. J. Biol. Chem. 278(37), 34998–35015 (2003)
3. Hua, S., Sun, Z.: Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17(8), 721–728 (2001)
4. Shen, H.B., Yang, J., Chou, K.C.: Methodology development for predicting subcellular localization and other attributes of proteins. Expert Rev. Proteomics 4(4), 453–463 (2007)
5. King, B.R., Guda, C.: ngLOC: an n-gram-based Bayesian method for estimating the subcellular proteomes of eukaryotes. Genome Biol. 8(5), R68 (2007)
6. Bulashevska, A., Eils, R.: Predicting protein subcellular locations using hierarchical ensemble of Bayesian classifiers based on Markov chains. BMC Bioinformatics 7, 298 (2006)
7. Mondal, A.M., Hu, J.: NetLoc: Network Based Protein Localization Prediction Using Protein-Protein Interaction and Co-expression Networks. In: IEEE International Conference on Bioinformatics & Biomedicine (BIBM 2010), Hong Kong (2010)

8. Lee, K., et al.: Protein networks markedly improve prediction of subcellular localization in multiple eukaryotic species. Nucleic Acids Res. 36(20), e136 (2008)
9. Mintz-Oron, S., et al.: Network-based prediction of metabolic enzymes' subcellular localization. Bioinformatics 25(12), i247–i252 (2009)
10. Mondal, A.M., Hu, J.: Network Based Prediction of Protein Localization Using Diffusion Kernel. International Journal of Data Mining and Bioinformatics (2011) (in press)
11. Mondal, A.M., Lin, J., Hu, J.: Network Based Subcellular Localization Prediction for Multi-Label Proteins. In: BIBM-International Workshop on Biomolecular Network Analysis (IWBNA) (2011)
12. Mondal, A.M., Hu, J.: Protein Localization by Integrating Multiple Protein Correlation Networks. In: The 2012 International Conference on Bioinformatics & Computational Biology (BIOCOMP 2012), Las Vegas, USA (2012)
13. von Mering, C., et al.: STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 33(database issue), D433–D437 (2005)
14. Lee, H., et al.: Diffusion kernel-based logistic regression models for protein function prediction. OMICS 10(1), 40–55 (2006)
15. Stark, C., et al.: BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 34(database issue), D535–D539 (2006)
16. Spellman, P.T., et al.: Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell. 9(12), 3273–3297 (1998)

An Evolutionary Approach for Analysing the Effect of Interaction Site Structural Features on Protein- Protein Complex Formation

Archana Chowdhury[1], Pratyusha Rakshit[1],
Amit Konar[1], and Ramadoss Janarthanan[2]

[1] Department of ETCE,Jadavpur University, Kolkata, India
[2] Department of CSE,TJS Engg. College, Chennai, India
{chowdhuryarchana,pratyushar1}@gmail.com, konaramit@yahoo.co.in,
srmjana_73@yahoo.com

Abstract. Protein-protein complexes that dissociate and associate readily,often depending on the physiological condition or environment,play an important role in many biological processes.The impact of the features responsible for protein complex formation is not uniform. In this paper we have tried to rank the features required for stable protein-protein complex formation.We have employed Artificial Bee Colony with Temporal Difference Q learning algorithm to assign weights to the various atomic structure features. Experiments with data provide evidence that such an approach leads to improved clustering performance.

Keywords: protein-protein complex, interactome, Artificial Bee Colony algorithm, Temporal Difference Q learning, clustering.

1 Introduction

Most proteins form complexes to accomplish their biological functions [1]. Proteins perform and regulate many processes in the cell through interactions with other proteins. It has been estimated that 70% of proteins act through multi protein complexes in yeast [2].In order to predict an interaction site in situations where the binding partner is unknown, more emphasis is paid to properties which are observed in unbound structures which includes physicochemical parameters [3]and the evolutionary conservation of amino acid residues [4].Recent advances in highthroughput experimental methods for the identification of protein interactions have resulted in a large amount of diverse data that are some what incomplete and contradictory.Such experimental approaches studying protein interactomes have certain limitations that can be complemented by the computational methods for predicting protein interactions.

In this paper, we aim to analyze the effect of protein-protein interaction site structural features on clustering the protein-protein complex. For this task we have employed Artificial Bee Colony with Temporal Difference Q-Learning

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 656–661, 2013.

(ABC-TDQL)[5].It is shown that ABC-TDQL, inspired by global search capability of ABC [6]and from principles of reward and penalty of reinforcement learning,can give very promising results.

The rest of the paper is organized as follows: Section 2 give a brief idea about the structural features of protein-protein interaction site. Section 3 defines clustering problem in a formal language.Section 4 gives a brief description about the Artificial Bee Colony with Temporal Difference Q-Learning. Experiments and Results are provided in Section5.Section 6 concludes the paper.

2 The Structure of Protein-Protein Interaction Sites

The atomic structure of the recognition sites found in 63 protein-protein complexes of known three-dimensional structure is taken from [7].Sample dataset includes features such as: interface area of complex, number of hydrogen bond, number of water molecules,% interface area for non-polar atoms at interface,% interface area for polar atom,% interface area for charged atoms, number of interface atoms, packing density of buried atoms at interface with % interface area and packing density of atoms surrounded by solvent with % interface area.The dataset consists of protein protein complexs which belong to four groups of protease-inhibitor, antibody-antigen,enzyme-inhibitor and those that are involved in signal transduction.

3 Formulation of Problem

Problem Definition: Let $X_{N\times D} = \left\{\vec{X_1}, \vec{X_2}, \cdots, \vec{X_N}\right\}$ be a set of N interaction sites of protein-protein complexes, each having D features. A partitional clustering algorithm tries to find out a partition of K clusters,such that the similarity of the interaction sites in the same cluster is maximum and interaction sites from different clusters differ as far as possible. Similarity is evaluated using the Euclidean distance $d(\vec{X_i}, \vec{X_j})$.

Clustering Validity Index (CS Measure): Let $\vec{m_i}$ be the centroid of i-th cluster of protein-protein complexes. The CS measure [8]is then defined as

$$CS(K) = \frac{\frac{1}{k}\sum_{i=1}^{k}\left[\frac{1}{N_i}\sum_{x_i\in C_i} max_{x_q\in C_i}\left\{d(X_i, X_q)\right\}\right]}{\frac{1}{k}\sum_{i=1}^{k}\left[min_{j\in k, j\neq i}\left\{d(m_i, m_j)\right\}\right]} \tag{1}$$

CS measure is a function of ratio of sum of within-cluster-scatter to between cluster separation. It tries to find out clusters that have minimum within-cluster scatter (i.e. compact) and maximum between-cluster separation (i.e. well-separated)

Solution Representation and Fitness Function Evaluation: In the proposed method, for D dimensional interaction sites of N protein-protein complexes,a solution is a vector of real numbers of dimension $D + D \times K$ where K

is the number of clusters. The first D entries of $\vec{Z_i}$ represent the weights of the features which are randomly initialized and the remaining entries are reserved for K cluster centers, each D dimensional.Representation of the solution $\vec{Z_i}$ is shown in the following figure with its fitness function given in (2).

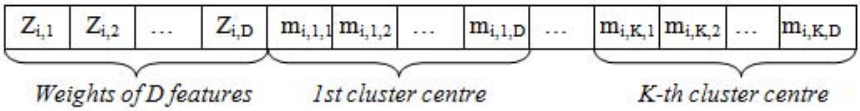


Fig. 1. Representation of a solution for ABC-TDQL-based clustering

$$f(\vec{Z_i}) = 1/CS_i(K) \tag{2}$$

4 Artificial Bee Colony with Temporal Difference Q-Learning (ABC-TDQL)

The ABC-TDQL [5] includes an ABC (with NP employed and NP onlooker bees) for global exploration and a TDQL for adaptive selection of scale factors for the individual members of the ABC using the Q-table. The row indices of the Q-table represent states $S_1, S_2, \cdots, S_{NP}$ of the population (based on fitness measure) obtained from the last update of the Q-table. The column indices correspond to uniformly quantized values of the scale factors $F_1, F_2, \cdots, F_{10}$ to be used in ABC. The steps of ABC-TDQL are given below.

1. Initialization: ABC-TDQL starts by initializing a population of NP, D-dimensional food sources (solutions)$\vec{Z_i}$ (G) at generation G=0 within the prescribed bounds and their fitness values (nectar amount)$fit(\vec{Z_i}\ (G))$ are evaluated for $i = [1, NP]$. The entries for the Q-table are initialized with small values.

2. Adaptive Selection of Parameters for Employed Bee Phase: The scale factor $F = F_j$ is randomly selected for individual food source from the meme pool satisfying(3)with r as a random number between (0,1).

$$\frac{\sum_{m=1}^{j-1} Q(S_i, 10F_m)}{\sum_{n=1}^{10} Q(S_i, 10F_n)} < r \leq \frac{\sum_{m=1}^{j} Q(S_i, 10F_m)}{\sum_{n=1}^{10} Q(S_i, 10F_n)} \tag{3}$$

3. Employed Bee Phase: An employed bee produces a modification $\vec{Z_i'}\ (G) = \left[Z_{i,1}(G), Z_{i,2}(G), \cdots, Z'_{i,j}(G), \cdots, Z_{i,D}(G)\right]$ on the position in her memory$\vec{Z_i}$ $(G) = [Z_{i,1}(G), Z_{i,2}(G), \cdots, Z_{i,j}(G), \cdots, Z_{i,D}(G)]$ and tests $fit(\vec{Z_i'}\ (G))$. $Z'_{i,j}(G)$ is computed using (4):

$$Z'_{i,j}(G) = Z_{i,j}(G) + 2 \times (F - 0.5) \times (Z_{i,j}(G) - Z_{k,j}(G)) \tag{4}$$

Here F is the scale factor in [0, 1] adaptively selected from the meme pool in the step 2, j and k are randomly selected such that $j \in [1, D]$, $k \in [1, NP]$, $k \neq i$. The bee replaces $Z_i(G)$ by $Z_i'(G)$ if $fit(\vec{Z_i'}\ (G))$>$fit(\vec{Z_i}\ (G))$.

4. Ranking of the Members: The food sources are sorted in descending order of probability of selection by onlooker bee as in (5).

$$prob(i) = fit(\vec{Z}_i\ (G))/\sum_{j=1}^{NP} fit(\vec{Z}_j\ (G)) \quad ,\forall i \tag{5}$$

5. Q-table Updating: Let a member at state S_i on selection of F_j moves to a new state S_k. Then $Q(S_i, 10F_j)$ will be updated following (7) with α and γ as the learning rate and discount factor respectively and

$$reward(S_i, 10F_j) = \left\{ \begin{array}{c} fit(\vec{Z_i'}(G)) - fit(\vec{Z_i}(G)),\ \ if\ fit(\vec{Z_i'}(G)) > fit(\vec{Z_i}(G)) \\ -K\ (howeversmall),\quad otherwise \end{array} \right\} \tag{6}$$

$$Q(S_i, 10F_j) = (1-\alpha)Q(S_i, 10F_j) + \alpha(reward(S_i, 10F_j) + \gamma max_{F'} Q(S_k, 10F')) \tag{7}$$

6. Onlooker Bee Phase: Every onlooker bee probabilistically selects a food source depending on the probability value as stated in (5). Next steps 2 to 5 are repeated.

7. Scout Bee Phase: The abandoned food source is reinitialized randomly by the scout.

8. Convergence: After each evolution, steps are repeated until the condition for convergence is satisfied.

5 Experiment and Results

ABC-TDQL algorithm is run for 500 generations with population size of 50.The proposed algorithm is compared with other algorithms such as DE-TDQL, SaDE, DE/current-to-best/1, PSO and GA-based clustering method to evaluate the efficiency in assigning weights to the features.Table 1, represents the mean and the standard deviation (within parentheses) of final CS value, the intracluster distance, and the intercluster distance obtained for 50 independent runs of each of the algorithms.Table 2 represents the results comparing speed of various algorithms in terms of mean and standard deviation (within parenthesis) of the number of FEs, the CS cutoff value(0.10), the final intracluster distance, and the final intercluster distance over 50 independent runs for each algorithm. As the nominal partitions of the dataset is known, the mean number of misclassified data points for different features is calculated and presented in Table 3.

Table 4 represents various features arranged according to their decreasing weights obtained using the proposed algorithm.The entries in Table 4 are relevant and supported by the fact that the most important feature which helps to form stable protein- protein complexes include hydrophobic effect, which is gained from the surfaces buried in the recognition sites i.e. interface area.Then next is the contribution by non polar atoms, as buried atoms are more non-polar, 63% on average, in comparison to other interface atoms which remain partly accessible.

Table 1. Final solution

Algo-rithm	CS Measure	Intra-clu.Dist	Inter-clu.Dist
ABC-TDQL	0.133770 (0.01365)	1.034500 (0.01054)	7.881300 (0.01388)
DE-TDQL	0.263210 (0.02642)	1.197500 (0.01210)	7.918300 (0.01054)
SaDE	0.409070 (0.04092)	1.900700 (0.01908)	6.516400 (0.02642)
DE/current-to-best/1	0.609710 (0.06154)	4.248900 (0.04252)	4.906400 (0.04344)
PSO	0.814950 (0.08207)	5.620500 (0.05624)	3.721800 (0.05824)
GA	0.947700 (0.09491)	7.014900 (0.07139)	1.337700 (0.07374)

Table 2. Speed of algorithms

Algo-rithm	No. of FEs	Intra-clu. Dist	Inter-clu. Dist
ABC-TDQL	144321.29 (14.5320)	1.878600 (0.01932)	9.227900 (0.01711)
DE-TDQL	144321.29 (14.5320)	1.878600 (0.01932)	9.227900 (0.01711)
SaDE	268983.67 (27.5100)	3.663000 (0.03716)	7.638800 (0.03709)
DE/current-to-best/1	457415.26 (46.9980)	5.207600 (0.05220)	6.223300 (0.05159)
PSO	650836.54 (65.1550)	6.661200 (0.06700)	4.542400 (0.07000)
GA	843198.42 (84.4520)	8.233900 (0.08280)	2.417200 (0.08673)

Table 3. Mean classification error over nominal partition and standard deviation

ABC-TDQL	DE-TDQL	SaDE	DE/current-to-best/1	PSO	GA	Stat. Sig.
1.55(0.01)	1.61(0.01)	2.28(0.03)	4.98(0.05)	6.34(0.06)	8.06(0.08)	+

Table 4. Sorted Features

Index	Sorted Features
f01	Interface Area
f03	Number of Hydrogen Bonds
f04	% Interface Area of Polar Atoms
f05	V/V0
f06	% Interface Area for V/V0
f07	V'/V0
f08	% Interface Area for V'/V0
f09	Number of interface atoms
f10	Number of Water Molecules
f11	% Interface Area of Charged Atoms

Table 5. Statistical comparision

Classifier - algorithm	n_{01}	n_{10}	Zj	Comment
DE-TDQL	19	28	1.3617	Accepted
SaDE	18	30	2.5208	Accepted
DE/current-to-best/1	12	40	14.019	Rejected Rejected
PSO	9	45	22.685	Rejected
GA	5	47	32.326	Rejected

On average, neutral polar groups contribute 29% to the interface area. Thus, the fraction contributed by non polar groups is higher than the fraction contributed by neutral polar groups, whereas charged groups contribution is the least. Then comes electrostatic energy from the hydrogen bonds. For interface atoms that are buried in 63 complexes, which represents one-third of all interface atoms, the packing density derived from the Voronoi volume is within 7% of that of the protein interior. Since two-thirds of the interface atoms have non-zero solvent accessibility, their volume cannot be calculated in the absence of information of the structure of the solvent molecules with which they are in contact hence their contribution to stable complex formation is very less. McNemer's test is applied to determine the performance of two algorithms used for clustering of data points and is expressed as follows :

$$Z = (|n_{01} - n_{10}| - 1)^2/(n_{01} + n_{10}) \quad (8)$$

where n_{01} and n_{10} give the number of samples misclassified by one algorithm and not by the other with which comparision is done and vice versa.Table 5 represents the Z and the null hypothesis is rejected if $Z > \chi^2_{1,0.95} = 3.841459$, which indicates that the probability of the null hypothesis is correct only to a level of 5%. It is apparent from Table 5 that ABC-TDQL-based clustering technique has outperformed most of its competitors in inferring clusters of the data points except DE-TDQL and SaDE.

6 Conclusion

This paper proposes a novel technique for selecting features based on a clustering algorithm. The observations suggest that the structural features of the recognition sites, if found in other systems where proteins exist as stable independently, will interact to form specific complexes without major conformational changes, since the conformation of complexes considered for observation were almost same. Hence further research may involve investigating the recognition site features of complexes that involve large conformational change.

References

1. Gavin, A.C., et al.: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415(6868), 141–147 (2002)
2. Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al.: Functional organization of the yeast proteome by systematic analysis of proteincomplexes. Nature 415, 141–147 (2002)
3. Janin, J., Chothia, C.: The structure of protein-protein recognition sites. J. Biol. Chem. 265, 16027–16030 (1990)
4. Lichtarge, O., Sowa, M.E.: Evolutionary predictions of binding surfaces and interactions. Curr. Opin. Struct. Biol. 12, 21–27 (2002)
5. Rakshit, P., Konar, A., Das, S., Nagar, A.K.: ABC-TDQL: an adap-tive memetic algorithm. In: 2013 IEEE Symposium Series on Computational Intelligence (accepted, to be published 2013)
6. Bhattacharjee, P., Rakshit, P., Goswami, I., Konar, A., Nagar, A.K.: Multi-robot path-planning using artificial bee colony optimization algorithm. In: NaBIC 2011, pp. 219–224 (2011)
7. Lo Conte, L., Chothia, C., Janin, J.: The Atomic Structure of Protein-Protein Recognition Sites. J. Mol. Biol. 285, 2177–2198 (1999)
8. Chou, C.H., Su, M.C., Lai, E.: A new cluster validity measure and its application to image compression. Pattern Anal. Appl. 7(2), 205–220 (2004)

Random Weighting through Linear Programming into Intracellular Transporters of Rice Metabolic Network

Rahul Shaw and Sudip Kundu

Department of Biophysics, Molecular Biology and Bioinformatics,
University of Calcutta, Kolkata 700 009, India
rahul.nutron@gmail.com, skbmbg@caluniv.ac.in
www.caluniv.ac.in

Abstract. An eukaryotic cell has different compartments which are specific to different biological activities. The total cellular metabolism is also compartmentalized. The intracellular transporters within a cell are responsible to transport some of the metabolites of one compartment to other. We formulate a model to understand the utility of different transporters. Here, we have taken a partially compartmentalized genome scale metabolic model of rice (*Oryza sativa*). Depending on the gene-expression, the transporters available to transport the metabolites from one compartment to other would change. We study the effect of transporter's capacity on the overall metabolism. We find that depending on the effectiveness of transporters, the photon demand for a rice leaf's biochemical machinery to synthesize the necessary biomass from inorganic nutrients, changes upto three fold. We also observe, interactions of mitochondrial and chloroplastid reactions are associated with this change.

Keywords: Random Weighting, Linear Programming, Metabolic Network, Genome Scale Model, Intracellular Transporter.

1 Introduction

Understanding the plant's cellular metabolism using Flux Balance Analysis (FBA) [1] is one of the active topic in systems biology. FBA is a mathematical approach, mainly based on linear programing, to analyze the flow of metabolites in a metabolic network. The genome scale metabolic model of plants are available [2, 3] and it has been demonstrated that FBA can be used to understand different cellular behaviors like the pathways involved in biomass production and also the change of pathways under varying conditions. In the FBA, one wants to optimize an objective function. In metabolic model, several optimization criterion including maximizing biomass composition or minimizing the total reaction flux have been used.

The human mitochondrial metabolic model has been analyzed for three different objectives using the Pareto optimality criteria and this also uses different

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 662–667, 2013.

weights [4]. Here, we have utilized the variation of weight in objective function to mimic the different cellular behavior under perturbed conditions.

Random weighting is used to develop genetic algorithm for bi-level mixed linear integer programming [5]. Multiobjective optimization using NISE method for conflicting objectives like maximizing succinic acid production versus maximizing biomass production for an in silico model of E Coli also uses different weights in each iteration to maximize each objective [6].

Unlike prokaryotes, eukaryotic cells have more complex structures with different compartments for specific tasks. Membrane bounded compartments form a diffusion barrier and prevent uncontrolled exchange of intermediates [7]. While weighting to non-transporters is related to mimicking the gene expression of enzymes, weighting to transporters is used to capture the capacity of transporting the metabolites from one compartment to other. Gene expression of a cell vary under different conditions and these conditions demands different transport mechanism in intracellular level. Thus, we are mimicking the possible active metabolic states of the cell under different conditions, in specific, here we determine the effect of intracellular transporter's transport capacity on the overall metabolism.

2 Materials and Methods

2.1 A Genome Scale Model of Rice

A Genome Scale Model (GSM) of rice (*Oryza sativa*), developed in our group [3] is used to reconstruct the network with different weight among Intracellular Transports (ICT). The constraints are same as used by (Poolman, 2013) [3]. This model has 9 mitochondrial and 14 chloroplastidial intracellular transporters. Here, all intracellular transporters are randomly weighted from 0 (no cost) to 1000 (extreme costly) except photon transporter in chloroplast which is free to use any value. Here, the flux through ATPase is not fixed, however it has a lower cut-off of 0.1. The biomass is fixed in the proportions as experimentally observed and other details are same as described in [3].

2.2 Linear Programming

Linear programming is a well known mathematical method used to achieve best possible solution among several options [8, 9]. Either we can minimize or maximize our objective such as lower flux (low cost) or higher biomass production (high gain) in plant system with some given constraints. Here, metabolic network model of rice is used to form a stoichiometric matrix

$$\mathbf{S}_{m\times n} \tag{1}$$

where m is number of metabolites and n is number of reactions. All the reactions in model is mass balanced; if $\mathbf{v}$ is the flux vector of reactions, then at steady state

$$\mathbf{S}.\mathbf{v} = 0 \tag{2}$$

$$l \leq \mathbf{v} \leq u \tag{3}$$

where l and u correspond to lower and upper bounds of the fluxes, respectively. Objective function in FBA is

$$Z = \mathbf{w}.\mathbf{v} \tag{4}$$

where $\mathbf{w}$ is the weight vector of reactions. We vary $\mathbf{w}$ to mimic the different transport capacity of the transporters. Here, we choose minimization of Z which corresponds to minimization of total reaction fluxes, *i.e.*, optimizing the cellular economy.

2.3 Algorithm

1. Load Model *//From a ScrumPy model file*
2. Build Linear Programming (LP) Object *//Include model in glpk-lp*
3. Set Constraints of LP *//Lower (l) and upper (u) bound of fluxes*

$$\begin{cases} l \leftarrow 0, u \leftarrow +\text{AV for irreversible reaction} \\ l \leftarrow -AV, u \leftarrow +\text{AV for reversible reaction} \end{cases} \quad \textit{//AV is Any Value of fluxes}$$

4. Set Fixed Flux of Biomass *//Flux of all biomass precursors are fixed in desired proportion*
5. Set Objective to Minimize Z *//Minimize total reaction fluxes*
6. T = List of Intracellular Transporters *//Includes 23 intracellular transporters*
   for $i = 1$ to cardinality of T, do
   $\mathbf{w}_{\mathrm{T}_i} \leftarrow$ A random integer in [0,1000] *//uniform integer distribution*
   end for *//Random weight assignment on all ICT is complete*
7. Solve LP Using Simplex Method *//to get flux distributions*
8. Get Solution *//Solution space of reactions with their respective fluxes*
9. End

The algorithm is shown only for a single iteration.

2.4 Package Used for Analysis

ScrumPy - is a Metabolic Modelling Tool implemented by Python programming language [10] and all computations are done using this.

3 Results and Discussion

We repeat the algorithm for 1000 times and accumulate the solution for each weight-set on ICT. These solutions allow us to predict the results described below.

3.1 General Responses

Among 1733 reactions 365 were used and 189 are always active (except biomass precursors which are fixed to active all the time). It indicates that these 189 reactions are essential for cell in any condition. We observe that O_2 and CO_2 transport in chloroplast and pyruvate and AlphaKG transport in mitochondria should remain active irrespective of any cost imposed on them. So they are essential for carrying metabolites to and from compartments which are important for plant survival and biomass production.

3.2 Photon Uses Efficiency

Minimum and maximum photon fluxes are 0.32 and 1.08, respectively. Figure 1 shows that maximum number of combinations of ICT weights use low photon while some of the combinations use high photon. We further observe that the higher amount of photon and thus energy is needed for some combinations of weight factors. Observing the weights we could say which of the intracellular transporters are favorable in different conditions. It is observed that the plant can synthesize the biomass with comparatively lower amount of photon when the chloroplastidial PGA and MalOxAc transporters are highly active.

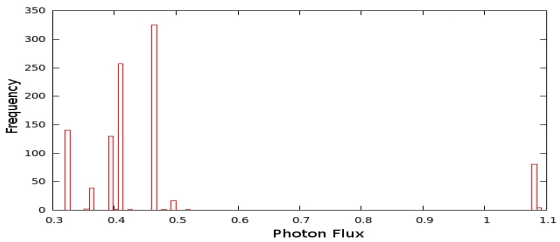


Fig. 1. A histogram of photon use. Photon uses can increase up to three times depending on the usability of ICT. Low photon use occurs for more than 85% of the combinations.

3.3 Observation of Compartmental Activity

There are several modes of TCA cycle throughout the solution, some are conventional cyclic and others are truncated or non-cyclic (Figure 2). Being one of the main energy production mechanism in aerobic organisms, these different modes reflects different energy demand conditions in cell. Cyclic mode establishes when demand for ATP is high or alternative sources are not producing sufficient ATP. When demand for ATP is low or alternative ATP sources are sufficient then TCA cycle becomes truncated. Some of the operative modes of TCA cycle are already reported in [2, 11].

When the photon flux is low, more ATP is generated by the mitochondrial Electron Transport Chain (ETC) and/or by full TCA cycle to achieve sufficient energy requirement. When sufficient ATP is generated by light reactions in the chloroplast, generation of extra ATP is prevented by maintaining low flux through Complex V (shown in the scatter plot in Figure 2) or controlling ETC and using different non cyclic modes. So generation and utilization of energy source is a controlled process and TCA modes are major controller of the same. At the same time, our results indicate that there exists a relationship of chloroplast and mitochondrial metabolism.

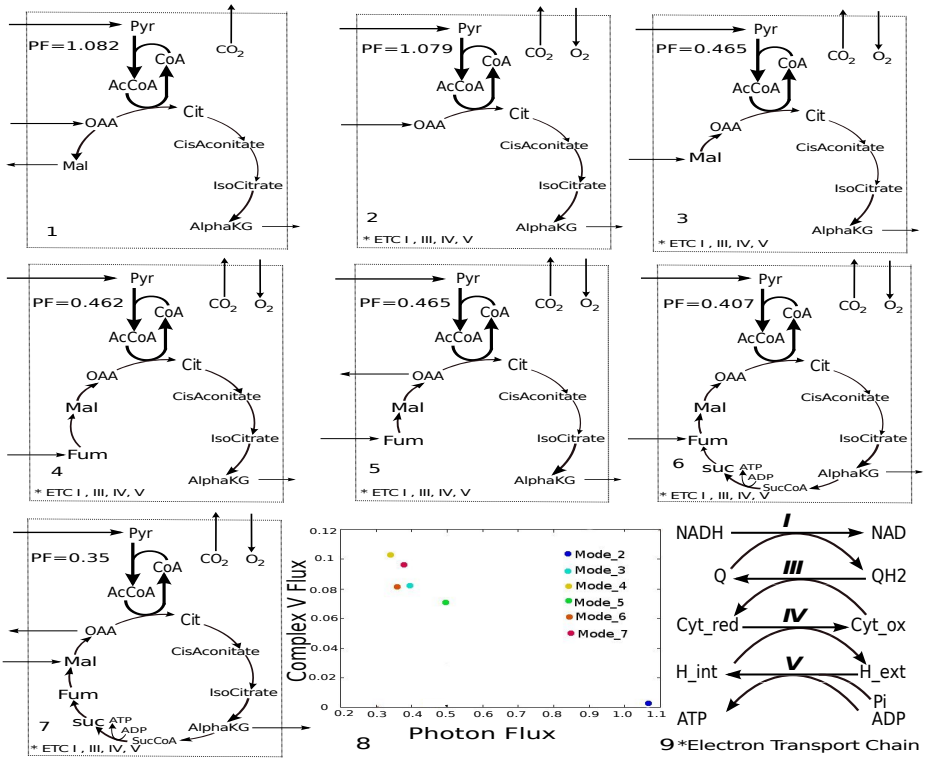


Fig. 2. Modes of TCA Cycle. PF is Photon Flux, ETC is Electron Transport Chain, 1: Non-cyclic mode which doesn't require ETC, 2,3,4,5: Non-cyclic modes using ETC, 6,7: Cyclic Modes, 8: Scatter Plot of Photon Flux Vs Complex V Flux, 9: Electron Transport Chain which is present in modes 2,3,4,5,6,7. Incoming arrow into the dotted box indicates metabolite going into mitochondria and vice versa.

4 Conclusions

Here, we have identified that i) the flux through the chloroplast transporters control the demand of photon to synthesize biomass ii) depending on the transporter's transport capacity, different modes of TCA cycle become active iii) there

exists an interaction of chloroplast and mitochondrial metabolism. Finally, this method can be used to mimic the effect of variations of enzymatic gene expressions in cellular metabolism.

Abbreviations Used in Figures and Elsewhere:
CoA:Coenzyme A; Pyr:Pyruvate; IsoCitrate:isocitrate; AlphaKG:alpha ketoglutarate; SucCoA:succinyl-CoA; Fum:Fumarate; Mal:Malate; OAA:oxaloacetate; AcCoA:Acetyl-CoA; CisAconitate:cis-aconitate; suc:succinate; Cit:citrate; Cyt _red:cytochrome c reductase; Cyt_ox:cytochrome c oxidase; Q:ubiquinone; QH2: ubiquinol; _ext: external; _int: internal; PGA: 3-phosphoglycerate; MalOxAc: malate oxaloacetate; PEP: phosphoenolpyruvate; G6P: glucose 6-phosphate, ATPase: ATP demand.

Acknowledgement. RS thanks Distributed Information Center (DIC), University of Calcutta funded by DBT, Govt. of India for his fellowship.

References

1. Edwards, J.S., Covert, M., Palsson, B.O.: Metabolic Modeling of Microbes: the Flux Balance Approach. Environmental Microbiology 3, 133–140 (2002)
2. Poolman, M.G., Miguet, L., Sweetlove, L.J., Fell, D.A.: A Genome-Scale Metabolic Model of Arabidopsis and Some of Its Properties. Plant Physiology 151, 1570–1581 (2009)
3. Poolman, M.G., Kundu, S., Shaw, R., Fell, D.A.: Responses to Light Intensity in a GenomeScale Model of Rice Metabolism. Plant Physiology 162, 1060–1072 (2013)
4. Vo, T.D., Greenberg, H.J., Palsson, B.O.: Reconstruction and Functional Characterization of the Human Mitochondrial Metabolic Network Based on Proteomic and Biochemical Data. The Jou. of Biol. Chem. 279, 39532–39540 (2004)
5. Zou, G., Jia, L., Zou, J.: Random-weight Based Genetic Algorithm for Multiobjective Bilevel Mixed Linear Integer Programming. In: IEEE ICNC (2012)
6. Oh, Y.G., Lee, D.Y., Lee, S.Y., Park, S.: Multiobjective Flux Balancing Using the NISE Method for Metabolic Network Analysis. Biotechnol. Prog. 25, 999–1008 (2009)
7. Linka, N., Weber, A.: Intracellular Metabolite Transporters in Plants. Molecular Plant 3, 21–53 (2010)
8. Zimmermann, H.J.: Fuzzy programming and linear programming with several objective functions. Fuzzy Sets and Systems 1, 45–55 (1978)
9. Karmarkar, N.: A new polynomial-time algorithm for linear programming. In: Proceeding STOC 1984, Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing, pp. 302–311 (1984)
10. Poolman, M.G.: ScrumPy: metabolic modelling with Python. IEE Proc. Syst. Biol. 153, 375–378 (2006)
11. Sweetlove, L.J., Beard, K.F.M., Nunes-Nesi, A., Fernie, A.R., Ratcliffe, G.: Not just a circle: flux modes in the plant TCA cycle. Trends in Plant Science 15, 462–470 (2010)

Fuzzy SVM with a Novel Membership Function for Prediction of Protein-Protein Interaction Sites in *Homo sapiens*

Brijesh Kumar Sriwastava[1], Subhadip Basu[2,*], and Ujjwal Maulik[2,*]

[1] Department of Computer Science and Engineering,
Government College of Engineering and Leather Technology, Kolkata-700098, India
{sriwastavabrijesh@yahoo.co.in}
[2] Department of Computer Science and Engineering, Jadavpur University,
Kolkata – 700032, India
{subhadip,umaulik}@cse.jdvu.ac.in

Abstract. Predicting residues that participate in protein–protein interactions (PPI) helps to identify the amino acids located at the interface. In this work, experimentally verified 3-D structures of protein complexes are used for building the training model and subsequent prediction protein interactions from sequence information. Fuzzy SVM (F-SVM), which is developed on top of the classical SVM, is an effective method to solve this problem and we demonstrate that the performance of the SVM can further be improved with the use of a custom-designed fuzzy membership function. We evaluate the performances of both SVM and F-SVM on the PPI database of the *Homo sapiens* organism and evaluate the statistical significance of F-SVM over classical SVM. To predict interaction sites in protein complexes, local composition of amino acids together with their physico-chemical characteristics are used. The F-SVM based residues prediction method exploits the membership function for each pair sequence fragment and in all cases F-SVM improves the performances obtained by the corresponding SVM classifiers. The F-SVM performance on the test samples is measured by area under ROC curve (AUC) as 80.16% which is around 1.55% higher than the classical SVM classifier.

Keywords: Protein–protein interaction, Support vector machine, Fuzzy SVM.

1 Introduction

Protein-protein interactions (PPI) are at the core of the entire interaction system of any living cell, making them the central hubs or major mediators for virtually every bio-chemical process. Two major types of complexes are observed, namely homodimers and heterodimers, where homodimers mostly form permanent and highly optimized complexes, generally by aligning hydrophobic interfaces. In contrast, in the case of hetero-complexes, hydrophobicity is indistinguishable from the rest of the surface [1-3]. Jones and Thornton [4] suggested importance of differentiating between

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 668–673, 2013.

aforementioned types of complexes, when analysing their intermolecular interfaces. Summarizing, significant research was done in the area of protein-protein interactions, yet the problem of interaction sites prediction is still not fully understood.

Considering the inherent complexity of the problem, we have chosen to use SVM as the underlying classifier. However, the classical SVM algorithm is found to be inadequate to address the natural ambiguity in many datasets like the one used here for PPI site prediction. Therefore, we have used Fuzzy SVM (F-SVM) with a novel membership function to design the binary classification system for each pair sequence fragment to determine their interaction status.

Traditionally logic regression and neural networks technologies were used to deal with these noisy and ambiguous data sets. F-SVM, another alternative to work with noisy datasets, was first proposed in [5], where each sample is given a fuzzy membership that denotes the strength of belongingness of one data point towards one class.

Classical SVM implicitly uses the kernel function which often maps all training data from input space into a higher dimensional feature space. Fuzzy support vector machines (F-SVMs), Lin and Wang also worked in the same way, except that a membership value is associated with each training vector. Here, the membership value is multiplied into the penalty term to provide variable weighting [5].

In this paper we demonstrate that the performance of SVM algorithm can further be improved with the use of a custom-designed fuzzy membership function, for the PPI prediction problem. In this regard, we first discuss briefly the SVM classifier and then we highlight the design of the F-SVM classifier with the novel membership function for each pair sequence fragment to encode their interaction strength. Finally, we evaluate the performances of both SVM and F-SVM on the PPI databases of Homo sapiens.

2 Methods

2.1 Fuzzy Support Vector Machine Classifier

Support Vector Machine is an important machine learning technique proposed by Vapnik and co-workers [6]. In classical SVM, each sample is treated equally, *i.e.*, each input point is assigned to either one of the two classes. However, in some problems, some input points, such as the outliers, may not be exactly assigned to one of these two classes. In this context each point does not have the same meaning to the decision surface. In order to solve this problem, fuzzy membership of each input point may be introduced in such a way that different input points can make different contribution to the construction of decision surface. Suppose the training samples with associated fuzzy membership are $(x_1, y_1, s_1), (x_2, y_2, s_2), \ldots\ldots, (x_l, y_l, s_l)$, where each $x_i \epsilon R^N$ is a training sample, $y_i \epsilon \{+1, -1\}$ represents their class label and s_i is the fuzzy membership of point x_i which satisfies the condition $\sigma \leq s_i \leq 1$, $i = \{1,2,,\ldots.,l\}$ and $\sigma > 0$.

The fuzzy membership s_i is the attitude of the corresponding point x_i to belong one class and the parameter ξ_i is a measure of error in the SVM, the term $s_i\xi_i$ is a measure of error with different weighting to belong to that class. The problem of finding the optimal hyperplane can then be formulated as:

$$\begin{cases} \min & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{l} s_i \xi_i \\ y_i(w^T x_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0 & \forall\ i = 1,2,\ldots,l \end{cases} \quad (1)$$

where, C is a constant. It is important to note that if s_i is small then $s_i\xi_i$ also becomes small and effect of parameter ξ_i in (1) reduced. Therefore, corresponding point x_i have lesser impact on decision solution.

In F-SVM, greatest lower bound of α_i is zero which is same as in classical SVM. However the lowest upper bound for α_i is s_iC, which is not constant as in classical SVM. So feasible region of α_i dynamically depends on fuzzy membership value (s_i) of point x_i belonging to that class.

The training set $T = \{(x_1, y_1), (x_2, y_2), \ldots\ldots, (x_l, y_l)\}$, where $x_i \epsilon R^N$ belongs to one of the class $y_i \epsilon(+1, -1)$ for $i = 1,2, \ldots\ldots, l$. Subsequently a 2-dimensional matrix R is computed for distance of each vector to other vectors of set

$$R = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1l} \\ d_{21} & d_{22} & \cdots & d_{2l} \\ \vdots & \vdots & \vdots & \vdots \\ d_{l1} & d_{l2} & \cdots & d_{ll} \end{bmatrix} \tag{2}$$

From the matrix we find the maximum distance as $d = \max_{i,j=1,2,\ldots\ldots,j} d_{ij}$. Then average distance of all sample points is calculated as:

$$D = \frac{1}{C_l^2}\sum_{i=1}^{l-1}\sum_{j=i+1}^{l} d_{ij} = \frac{2}{l(l-1)}\sum_{i=1}^{l-1}\sum_{j=i+1}^{l} d_{ij} \tag{3}$$

Let the density of Sample Point be ρ_i and higher value of ρ_i signifies that more points are flanking to it and the corresponding to this points, the SVM have major role in classification, whereas the lower value of ρ_i means there are sparse points near to x_i and hence the role of SVM in classification is lesser. Then, we multiply the posterior probability with density ratio $\frac{\rho_i}{\rho}$ to reflect the possibility of the sample belongs to the class that constitutes and based on the posterior probability weighting membership function [7]. We need to understand the sample distribution position and can adjust samples in the role of classification which are near to the hyperplane. This ratio may be greater than 1 or less than 1 but when it less than 1, it signifies that point is outlier or noise point. Then fuzzy membership is defined as follows:

$$\mu(x_i) = P(\omega_j|x_i).\frac{\rho_i}{\rho} \quad , i = 1,2, \ldots\ldots, l \ \ and \ j = 1,2 \tag{4}$$

2.2 Fuzzy Membership Evaluation

In practice, we determine the membership function based on domain experience. We are working with 21 size window fragment (win_size) [8, 9] and we are considering a fragment to be positive if there are at least 2 interactions. The choice of two interactions is made to avoid possible noise, in the form of isolated interaction residue, in the positive dataset. Now maximum number of interactions in feature vector of size 21 is $21 \times 21 = 441$. But, every feature vector is not of equal strength with respect to number of interaction. So, it is better that each feature vector will not give equal contribution for training i.e. the feature vector having higher interaction strength is likely to be have more impact on training rather than lower strength feature vector. This is just due to the fact that, if there are more interactions in a fragment then more positive features are trained and gives better decision capability of choosing positive and negative class when the test would performed. This idea is used to give fuzzy membership strength to the each feature vector which is as below:

$$fs_i = \frac{num\_it}{(win\_size \times win\_size)} \qquad , i = 1,2, \dots \dots, l \tag{5}$$

where fs_i is feature strength of i^{th} vector, num_it is number of interaction in i^{th} feature vector and win_size is window size. Now, each feature vector has its own strength and we have taken average of all such vector and that is used to give weight among all vectors. This is separately done for $+ve$ set of feature vector and $-ve$ set of feature vector:

$$fp_i = \frac{fs_i}{\sum_{i=1}^{n_p} fs_i}, i = 1,2, \dots, n_p \quad \text{and} \quad fn_j = \frac{fs_j}{\sum_{j=1}^{n_n} fs_j}, j = 1,2, \dots, n_n \tag{6}$$

where, fp_i , fn_i are fuzzy membership values for $+ve$ and $-ve$ feature vectors respectively and n_p is number of $+ve$ feature vectors and n_n is that of $-ve$ feature vectors. Finally, we define the fuzzy membership based on nature of our problem is defined in eqn. (7) as:

$$\mu_i = \mu(x_i) = \begin{cases} P(+|x_i).\frac{\rho_i}{\rho}.fp_i\,, & y_i = +1 \ \ and \ \ , i = 1,2, \dots \dots, n_p \\ P(-|x_i).\frac{\rho_i}{\rho}.fn_i, & y_i = -1 \ \ and \ \ , i = 1,2, \dots \dots, n_n \end{cases} \tag{7}$$

The overall process of the fuzzy membership evaluation function is explained in Algorithm 1.

Algorithm 1. Fuzzy membership function evaluation

Input: We have the training set $= \{(x_1, y_1), (x_2, y_2), \dots \dots, (x_l, y_l)\}$, where $x_i \epsilon R^N$ belongs to one of the class for $y_i \epsilon(+1, -1)$ for $i = 1,2, \dots \dots, l$..

Output: The fuzzy membership μ_i , $i = 1,2, \dots \dots, l$.

Step 1 (a) Compute a 2-dimensional matrix R for distance of each vector to other vectors of set using (2).

(b) From the matrix R, find the maximum distance as:

$$d = \max_{i,j=1,2,\dots,j} d_{ij}$$

(c) Compute the average distance of all sample points using (3)

Note that the definition of sample density mainly uses the distance between the sample points of the original space. Because the specific form of $\sqrt{(x)}$ is unknown, the distance between the sample points in feature space can be obtained by kernel function $K(x, x')$

$$\begin{aligned} d\big(\phi(x), \phi(x')\big) = \|\phi(x) - \phi(x')\| &= \sqrt{\big(\phi(x) - \phi(x')\big)^2} \\ &= \sqrt{\phi(x)^2 - 2\big(\phi(x).\phi(x')\big) + \phi(x')^2} \\ &= \sqrt{K(x,x) - 2.K(x,x') + K(x',x')} \end{aligned}$$

The distance between any two samples points after mapping to the feature space can be found out through the above formulation.

Step 2 Compute the priori class probability $P\big(\omega_j\big)$, $j = 1,2$.

Step 3 Compute the class conditional probability $p\big(x_i|\omega_j\big)$.

Step 4 Compute the posterior probability $P\big(\omega_j|x_i\big)$.

Step 5 Compute the average density ρ and ρ_i for the sample set .

Step 6 Compute the fuzzy membership $\mu(x_i)$ using (4).

Step 7 Finally update the fuzzy membership μ_i using (5) to (7).

3 Results

For our analysis, we have used the Protein Data Bank (PDB) [10], and the Database of Interacting Proteins (DIP) [11], databases. The complete database, cross validation datasets and the source code for fuzzy SVM tool, developed under the current work is available freely to download for academic users from our website http://code.google.com/p/cmater-bioinfo/. The database involves 2008 positive interactions and 2408 negative interactions for *Homo sapiens* proteome [9]. It may be noted that the number of positive and negative interactions, considered in the experiment dataset for any proteome, are only a subset of all possible positive and negative interactions. This is done so, to limit the computational complexity of the training algorithm, during the multi-fold cross validation (CV) process. Each interacting or non-interacting residue fragments are represented using HQI8 amino acids indices [12] for both positive and negative data samples for the selected organism. Finally, we compare cross validation results with the fuzzy and classical SVM classifiers.

To analyze the performance of the developed technique, we have done a 10-fold cross validation experiment on the dataset, discussed above. In both the classical SVM and the fuzzy SVM classifier, we have used *polynomial* kernel function of degree 5 during experiments over the cross validation set. Ten cross validation experiment runs are marked as $run\#1, run\#2, \ldots, run\#10$. For each run of the experiment in both classical and fuzzy SVM training program, we vary three key kernel parameters $(c, \gamma \; and \; r)$ within a finite range. During any run of the cross validation experiment (run_i), the optimum set of kernel parameters are estimated as p_i [13, 14] and the best results in each run are reported in the results. The average cross validation performances using the classical SVM and fuzzy SVM classifiers on the three organisms are given in Table 1. In case of *Homo sapiens*, we see 1.55% of AUC improvement from classical SVM to F-SVM. As we know that MCC gives idea over the quality of binary classification and we have observed that there is significant improvement of MCC from classical SVM to F-SVM. The MCC gains of F-SVM over classical SVM classifier are 3.01% on *Homo sapiens* dataset (see Table 1). One of important classification parameter Sensitivity are also improved from classical SVM to F-SVM which are 2.14% on *Homo sapiens* dataset (see Table 1) respectively. However specificity shows lower rate of improvement whose respective gains are as 0.96% on *Homo sapiens* dataset (see Table 1). The test accuracy measure parameter, F-measure, is also improved by 1.55%, on *Homo sapiens* dataset (see Table 1).

Table 1. Comparison of average performances of 10 fold CV experiment over *Homo sapiens* data using classical SVM and fuzzy SVM classifiers

Classifier	Sensitivity	Specificity	MCC	F-measure	AUC
SVM	0.76495	0.80731	0.57285	0.76646	0.78613
F-SVM	0.78634	0.81687	0.60336	0.78392	0.80161

4 Conclusion

In the present work, we introduce the fuzzy SVM as a novel and accurate classifier for PPI site prediction with better performance than classical SVM classifier. We have

designed a new fuzzy membership function to give fuzzy value to each of the positive and negative fragment based on their interacting strength and with the help of the Bayesian formula also. We first transform the membership of F-SVM through posterior probability and weighted. It has been observed that through the simulation experiment in the same data set based on the posterior probability weighting membership of fuzzy support vector machine in this paper and it is better than classical SVM in classification AUC results. It shows a better performance over all the three organisms.

In this paper, we have worked with fuzzy classifier over Homo sapiens organism specific database. We would like to work it on other database by including more organisms in near future. Due to limitation of computing resources, all interactions could not be considered for CV experiment. Despite certain constraints, the current version of fuzzy SVM is observed to generate a steady and balanced prediction result over CV data set samples of the selected organisms. The fuzzy SVM classifier is also made available for free download in the public domain.

References

1. Korn, A., Burnett, R.: Distribution and complementarity of hydropathy in multi-subunit proteins. PROTEINS: Structure, Function, and Bioinformatics 9, 37–55 (1991)
2. Jones, S., Thornton, J.M.: Analysis of Protein-Protein Interaction Sites using Surface Patches. JMB 272, 121–132 (1997)
3. Lo Conte, L., Chothia, C., Janin, J.: The atomic structure of protein– protein recognition sites. J. Mol. Biol. 285, 2177–2198 (1999)
4. Jones, S., Thornton, J.M.: Principles of protein-protein interactions. Proceedings of the National Academy of Sciences 93(1), 13–20 (1996)
5. Lin, C.-F., Wang, S.-D.: Fuzzy Support Vector Machines. IEEE Transactions on Neural Networks 13(2) (2002)
6. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
7. Wei, Y., Wu, X.: A New Fuzzy SVM based on the Posterior Probability Weighting Membership. Journal of Computers 7(6), 1385–1392 (2012)
8. Sriwastava, B.K., Basu, S., Maulik, U., Plewczynski, D.: Prediction of E. coli Protein-Protein Interaction Sites Using Inter-Residue Distances and High-Quality-Index Features. In: Satapathy, S.C., Avadhani, P.S., Abraham, A. (eds.) Proceedings of the InConINDIA 2012. AISC, vol. 132, pp. 837–844. Springer, Heidelberg (2012)
9. Sriwastava, B.K., Basu, S., Maulik, U., et al.: PPIcons: identification of protein-protein interaction sites in selected organisms. Journal of Molecular Modeling, 1–12 (2013)
10. Berman, H., Westbrook, J., Feng, Z., et al.: The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000)
11. Salwinski, L., Miller, C.S., Smith, A.J., et al.: The Database of Interacting Proteins: 2004 update. Nucleic Acids Research 32, D449–D451 (2004)
12. Saha, I., Maulik, U., Bandyopadhyay, S., et al.: Fuzzy Clustering of Physicochemical and Biochemical Properties of Amino Acids. Amino Acids (2011)
13. Basu, S., Plewczynski, D.: AMS3.0: prediction of post-translational modifications. BMC Bioinformatics 11, 210 (2010)
14. Plewczynski, D., Basu, S., Saha, I.: AMS 4.0: consensus prediction of post-translational modifications in protein sequences. Amino Acids 43(2), 573–582 (2012)

A Consensus Approach for Identification of Protein-Protein Interaction Sites in *Homo Sapiens*

Brijesh K. Sriwastava[1], Subhadip Basu[2,*], Ujjwal Maulik[2,*], and Dariusz Plewczynski[3]

[1] Department of Computer Science and Engineering, Government College of Engineering and Leather Technology, Kolkata-700098, India
sriwastavabrijesh@yahoo.co.in
[2] Department of Computer Science and Engineering, Jadavpur University, Kolkata – 700032, India
subhadip@cse.jdvu.ac.in, umaulik@cse.jdvu.ac.in
[3] Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, 02-106 Warsaw, Poland
d.plewczynski@icm.edu.pl

Abstract. The physico-chemical properties of interaction interfaces have a crucial role in characterization of protein–protein interactions. Given the unbound structure of a protein and the fact that it forms a complex with another known protein, the objective of this work is to identify the residues that are involved in the interaction. We attempt to predict interaction sites in protein complexes using local composition of amino acids together with their physico-chemical characteristics. The local sequence segments are dissected from the protein sequences using sliding window of 21 amino acids. The list of LSSs is passed to the support vector machine (SVM) predictor, which identifies interacting residue pairs considering their inter-atom distances. Three different SVM predictors are designed that generate area under ROC curve (AUC), Recall and Precision optimized results. Finally a *3-star* consensus strategy is designed to analyze 33 hetero-complexes of the *Homo sapiens* organism. The consensus approach generates the AUC score of 0.7376, which is superior to the individual SVM classification results.

Keywords: protein-protein interactions, machine learning, support vector machine, consensus approach.

1 Introduction

Protein-protein interactions (PPI) are at the core of the entire interaction system of any living cell, making them the central hubs or major mediators for virtually every bio-chemical process. Therefore for a given protein, in order to understand its biological function, it is important to identify its likely interactions with other proteins. Detailed information of protein-protein interactions, metabolic and signal transduction networks improves our understanding of diseases, perturbation of healthy states or

* Corresponding authors.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 674–679, 2013.

processes, providing the theoretical basis for new therapeutic approaches, mutant engineering and design, high throughput screening for drug design [1]. Currently developed high-throughput experimental methods, such as Yeast two-hybrid, or mass spectrometry provided the global view of the whole interaction network for model organisms (interactome) [2, 3].

Summarizing, significant research was done in the area of protein-protein interactions, yet the problem of interaction sites prediction is still not fully understood. Major unresolved issues are, among others, linked with the problem of selection of biological and physico-chemical features crucial for protein-protein interactions [4]. Moreover, any single physico-chemical feature is not sufficient to distinguish interface and non-interface residues, the complex nonlinear combinations of features are needed to describe an interaction site.

The PPI prediction is not the balanced learning problem; therefore the optimal set of computational methods' parameters is not easy to obtain. To select the proper subset of descriptors, Saha et al. have applied the consensus fuzzy clustering technique [5] to extract high quality physico-chemical indices from the set of 544 indices provided by the AAindex1 database (http://www.genome.jp/aaindex/). Deng et al. [6] proposed ensemble learning method in order to overcome the misbalancing problem in PPI and effectively utilize a wide variety of features. He combined bootstrap sampling technique, SVM-based fusion classifiers and weighted voting strategy.

In view of the above facts, the goal of our paper is to predict the interacting residues for a pair of proteins given their unbound structures. The interface residues define the interaction site for those two proteins. More specifically, we attempt to predict interaction sites in protein complexes more accurately using selective high quality index physico-chemical features (HQI) extracted from AAindex1 dataset. We have used the sliding window algorithm, used in many earlier works [7, 8], with the length of 21 amino acids to select sets of local sequence segments for each protein, then identifying interacting residue pairs by considering their inter-atom distances.

We have trained our method on the dataset of interacting proteins for *Homo sapiens* and evaluated the PPI site prediction performance on unknown test samples using SVM classifier. Three different classification systems are designed using Recall, Precision and AUC optimized training strategies. Finally a 3-star consensus approach is used to further improve the prediction performance. The dataset design principles, feature selection and classifier design methodologies are described in details in the following section. The results section provides the performance evaluation metrics and analysis of the prediction results for the developed PPI software.

2 Methods

2.1 Training Dataset

For our analysis, we selected two major databases containing experimental information about protein-protein interactions, namely Protein Data Bank (PDB) [9], where one can find the three-dimensional structures of protein complexes, and Database of Interacting Proteins (DIP, http://dip.doe-mbi.ucla.edu/dip) [10], where the known interactions among protein pairs are stored.

In the case of PPI interactions for *Homo sapiens*, we have taken the database from our previous work as mentioned in [7]. The complete databases and source code are available freely to download for academic users from our website http://code.google.com/p/cmater-bioinfo/.

2.2 Representation of PPI Features

Here, we are working with interacting protein pairs (say, P_A and P_B) from our aforementioned training datasets. Let P_A and P_B has their be described by their own amino acid sequences as $a_1, a_2, \ldots, a_M$ and $b_1, b_2, \ldots, b_N$ respectively, where $a_i, b_j \in \{A, R, N, D, L, K, M, F, C, Q, E, G, H, I, P, S, T, W, Y, V\}, \forall\, i = 1\ to\ M\ and\ \forall j = 1\ to\ N$

In the next step, we compute inter-atom distances between P_A and P_B . Please note that we consider only the heavy atoms (as given in respective PDB entry) from each amino acid for this purpose. We define the distance measures as follows: $D_P(a_i, b_j) = \min\left(d_r(a_{ik}, b_{jl})\right), \forall k = 1\ to\ P\ and\ \forall l = 1\ to\ Q$

Where P and $Q$are number of heavy atoms in the residues a_i and b_j respectively and $d_r(a_{ik}, b_{jl}) =$ inter-atom Euclidean distances between the K^{th} heavy atom of a_i and l^{th} heavy atom of b_j. If $D_P(a_i, b_j) =$ is lower than 3.5 Å[11], then corresponding residue pair (a_i, b_j) corresponding belonging to the protein pair (P_A, Q_B) is said to be interacting, otherwise they are said to be non-interacting.

The protein sequences of hetero-complexes are therefore divided into multiple overlapping segments of sub-sequences, each consisting of 21 amino acids. Please note that the results from our current study strongly support selection of 21 window size, providing optimal results for protein-protein interaction prediction as tested on sample subsets of pairs of interacting proteins. For each pair of local sequence segments (LSS) from proteins P_A and P_B we consider all residues from $a_1, a_2, \ldots, a_{21}$ and $b_1, b_2, \ldots, b_{21}$ respectively, and check whether any of the residue pairs has$D_P(a_i, b_j) < 3.5\ \dot{A}$. If found, we annotate the given pair of sub-sequences (obtained from P_A and P_B respectively) as positive., i.e. confirmed interaction and extract HQI8 features for the 42 residues. The overlapping subsequences are then shifted, as a sliding window, to check for further interactions. In all cases, where two sub-sequences have no interacting residue pair, then such sub-sequence pair is said to be non-interacting, and we recognized it as negative training cases described by 336 dimensional vectors of features using also HQI8 features [5]. These positive and negative vectors are then used by the machine learning procedure to train the support vector machine algorithm, designed separately to produce optimal recall, precision and AUC (Area under ROC curve) scores.

3 Results

The current work, reported in this paper, involves 3488 positive interactions and 5470 negative interactions for *Homo sapiens* proteome proteome. It may be noted that the number of positive and negative interactions, considered in the training dataset for any proteome are only a subset of all possible positive and negative interactions. This is done so to limit the computational complexity of the training algorithm, during the

multi-fold Cross Validation (CV) process. Each interacting or non-interacting residue fragments are represented using HQI8 amino acid indices for both positive and negative data samples. We discuss here the training and testing prediction results for the aforementioned organism.

To analyze the performance of the developed technique, we have designed a two-stage evaluation strategy. In the first stage, the overall dataset is divided into two parts with the ratio 88:12 to define the CV set and the test set respectively. Then 10-fold cross validation is done with the CV set. In the second stage, the optimum network (chosen from the best of the 10 runs during the CV experiment) is selected to evaluate the performance over the independent test set. AUC optimized results of 10-fold CV experiment (average performance) and over the independent test set are shown in Table 1. It also shows the performance analysis over the *Homo sapiens* dataset during the consensus approach using three different optimization strategies. The supplementary data sheets with detailed experiments with various optimization strategies are available in the public domain at http://code.google.com/p/cmater-bioinfo/.

Table 1. Result of ARP 3-consensus and AUC optimization over *Homo sapiens* CV set and test set

Run	Accuracy	Recall	Precision	Specificity	AUC	MCC	Fmeasure
CVAverage	83.77995	0.745418	0.821748	0.896709	0.821064	0.655168	0.781616
Test Set	72.27191	0.721839	0.624254	0.72328	0.722559	0.436224	0.66951
1star	73.703	0.74023	0.640159	0.734993	0.737611	0.465726	0.686567
2star	72.2719	0.721839	0.624254	0.72328	0.722559	0.436224	0.66951
3star	70.0358	0.705747	0.597276	0.696925	0.701336	0.393919	0.646997

We compared our results with similar works reported previously in the literature. In the work of Wang et al.[12] position specific scoring matrices (PSSMs) were used along with evolutionary conservation score for 11 neighbour residues. They obtained 71.9% AUC, 68.6 % Sensitivity and 65.4% Specificity over their dataset of 113 pairs of interacting proteins. Nguyen et al. [13] used PSSMs and accessible surface areas (ASA) with 15 neighbour residue to get 74.9% AUC, 35.9% Sensitivity and 92.9% Specificity scores over 77 individual proteins collected from the Protein Data Bank. Both the above methods used SVM pattern classifier. Deng et al. [6] used an ensemble method with weighted voting strategy along with SVM approach and achieved 79.7% AUC, 76.7% Sensitivity and 63.1% Specificity over 54 hetero-complexes. Bordner and Abagyan[14] achieved 76% Accuracy, 57% Recall and 26% Precision over 1494 protein-protein interfaces, of which 518 were homodimers, 114 were heterodimers and 862 were multimers. Singh et al.[11] obtained 60% Sensitivity and 75% Specificity in their Struct2Net web server.

In comparison, for *Homo sapiens* test data, we have obtained we have obtained 73.76% AUC, 74.02% Sensitivity (or Recall) and 73.5% Specificity (see Table 1). We have also calculated the MCC for the organism *Homo sapiens* which is 46.57% (given in Table 1). The F-measure are also calculated which is 68.66% (given in Table 1). We have also added Table 2 for easy comparison of our work with the existing ones available in the literature.

We have used 3-consensus results over the *Homo sapiens* test dataset for the ARP optimized networks are shown in Tables 2. Here, we have observed that when we take

the result of 1-star, there is little more improvement of all classifier metrics from our last work [7]. Recall got 1.84% more improvement, Precision improved by 1.6%, specificity, AUC, MCC, Fmeasure are improved by 1.2%, 1.5%, 2.95%, 1.71% respectively from PPIcons result over the same aforementioned organism. We have also performed 3-consensus over three different optimized classifier AUC, Recall, Precision whose are shown in table S4, S5 and S6 respectively in supplementary sheet.

Table 2. Comparison of our current work with the existing techniques

	Methods	AUC	Sensitivity	Specificity
1	Wang et. al. [12]	0.71933	0.68640	0.65417
2	Nguyen et. al. [13]	0.74943	0.3598	0.92949
3	Deng et. al. [6]	0.79761	0.76765	0.63158
4	Borderner et. al. [14]	-	0.57	0.26
5	Singh et. al. [11]	-	0.6	0.75
6	PPIcons (Homosapiens)[7]	0.722559	0.721839	0.72328
7	Current Method (Homosapiens)	0.737611	0.74023	0.734993

4 Conclusion

In the present work, we introduce the PPI software as a novel and accurate tool for PPI site prediction, using only protein sequences. This prediction model allows annotating unknown interactions, enriching the biological knowledge about proteins' partners. The performance of our predictor is better than most of the methods discussed in this paper. Although the datasets used in different works are sometimes different, up to now the general performance scores from different publications are compared in evaluation of different *in silico* methods in PPI domain.

In this paper, due to limitation of computing resources, all interactions could not be considered for training. Despite certain constraints, the current version of this consensus based approach is observed to generate a steady and balanced prediction result. As evident from the discussion in the Results section, the performance of the consensus based program is found to be comparable or better than the state-of-the-art tools available today. For most of the existing predictors their performances are not balanced, producing high Sensitivity, yet low Specificity, or vice-versa. Avoiding such a biasing is often difficult in a complex binary classification problem. Considering that, the balanced prediction potential of our developed algorithm may be considered as a good statistical learning characteristic. In future we plan to incorporate a larger training/test datasets, incorporating more proteins from *Yeast*, *E. coli* and other organisms, for design of improved versions of this approach. Brainstorming consensus [15] or weighted Markov chain based rank aggression approach[16] , ensemble classifier [17, 18] and some better idea from [19] may be used in the future to achieve such an objective.

References

1. Chelliah, V., Chen, L., Blundell, T., Lovell, S.: Distinguishing structural and functional restraints in evolution inorder to identify interaction sites. Journal of Molecular Biology 342, 1487–1504 (2004)

2. Uetz, P., Giot, L., Cagney, G.: A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature 403, 623–627 (2000)
3. Yuen, H., Gruhler, A., Heilbut, A.: Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415, 180–183 (2002)
4. Chen, X., Jeong, J.: Sequence-based Prediction of Protein Interaction Sites with an Integrative Method. Bioinformatics 25(5), 585–591 (2009)
5. Saha, I., Maulik, U., Bandyopadhyay, S., Plewczynski, D.: Fuzzy Clustering of Physicochemical and Biochemical Properties of Amino Acids. Amino Acids (2011)
6. Deng, L., Guan, J., Dong, Q., Zhou, S.: Prediction of protein-protein interaction sites using an ensemble method. BMC Bioinformatics 10, 426 (2009)
7. Sriwastava, B.K., Basu, S., Maulik, U., Plewczynski, D.: PPIcons: identification of protein-protein interaction sites in selected organisms. Journal of Molecular Modeling (accepted for publication, 2013)
8. Sriwastava, B.K., Basu, S., Maulik, U., Plewczynski, D.: Prediction of E.coli Protein-Protein Interaction Sites Using Inter-Residue Distances and High-Quality-Index Features. In: Satapathy, S.C., Avadhani, P.S., Abraham, A. (eds.) Proceedings of the InConINDIA 2012. AISC, vol. 132, pp. 837–844. Springer, Heidelberg (2012)
9. Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., Bourne, P.: The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000)
10. Salwinski, L., Miller, C.S., Smith, A.J., Pettit, F.K., Bowie, J.U., Eisenberg, D.: The Database of Interacting Proteins: 2004 update. Nucleic Acids Research 32, D449–D451 (2004)
11. Singh, R., Park, D., Xu, J., Hosur, R., Berger, B.: Struct2Net: a web service to predict protein–protein interactions using a structure-based approach. Nucleic Acids Research 38, W508–W515 (2010)
12. Wang, B., Chen, P., Huang, D.-S., Lia, J.-J., Lokc, T.-M., Lyud, M.R.: Predicting protein interaction sites from residue spatial sequence profile and evolution rate. Federation of European Biochemical Societies Letters 580, 380–384 (2006)
13. Nguyen, M.N., Rajapakse, J.C.: Protein-Protein Interface Residue Prediction with SVM Using Evolutionary Profiles and Accessible Surface Areas. IEEE (2006)
14. Bordner, A.J., Abagyan, R.: Statistical Analysis and Prediction of Protein–Protein Interfaces. PROTEINS: Structure, Function, and Bioinformatics 60, 353–366 (2005)
15. Plewczynski, D.: Brainstorming: weighted voting prediction of inhibitors for protein targets. Journal of Molecular Modeling 17, 2133–2141 (2010)
16. Sengupta, D., Maulik, U., Bandyopadhyay, S.: Weighted Markov Chain Based Aggregation of Biomolecule Orderings. IEEE/ACM Transactions on Computational Biology And Bioinformatics 9, 924–933 (2012)
17. Maulik, U., Chakraborty, D.: A self-trained ensemble with semisupervised SVM: An application to pixel classification of remote sensing imagery. Pattern Recognition 44, 615–623 (2011)
18. Maulik, U., Mukhopadhyay, A., Bandyopadhyay, S.: Combining Pareto-optimal clusters using supervised learning for identifying co-expressed genes. BMC Bioinformatics 10, 27 (2009)
19. Maulik, U., Bandyopadhyay, S., Wang, J.T.: Computational Intelligence and Pattern Analysis in Biology Informatics. Wiley. com. (2011)

Gaussian Fuzzy Index (GFI) for Cluster Validation: Identification of High Quality Biologically Enriched Clusters of Genes and Selection of Some Possible Genes Mediating Lung Cancer

Anupam Ghosh[1] and Rajat K. De[2]

[1] Department of Computer Science and Engineering,
Netaji Subhash Engineering College, Kolkata, India
anupam.ghosh@rediffmail.com
[2] Department of Machine Intelligence Unit, Indian Statistical Institute,
Kolkata, India
rajat@isical.ac.in

Abstract. In this article, we propose an index, called Gaussian Fuzzy-index (GFI), based on the notion of fuzzy set theory, for validating the clusters obtained by a clustering algorithm. This index is then used to identify some genes that have altered quite significantly from normal stage to diseased stage with respect to their expression patterns. Thus we can predict some possible disease mediating genes from microarray gene expression data. The methodology has been demonstrated on the gene expression data set dealing with human lung cancer. The performance of GFI is compared with 8 existing cluster validity indices. The results are appropriately validated using biochemical pathways. We have also implemented different cluster validity indices to demonstrate superior capability of GFI over the others.

1 Introduction

With the advent of high throughput technology, huge amount of data in various fields, including the field of molecular biology, is being generated/updated. In order to extract useful information from this huge data resource, people are finding interest in developing and using data exploration techniques. Clustering is a tool, in this regard, in finding natural groups of similar data, under unsupervision. The quality of the clusters obtained by an algorithm needs to be adjudicated or validated. Thus, cluster validation is major and challenging task [1]. There exist several cluster validity indices in literature [2]. Some of them are Dunn index (DI) [3], Davis-Bouldin index (DBI) [4], Silhouette index (SLI) [5], C-index (CI) [6], Goodman-Kruskal index (GKI) [7], Isolation index (II) [8] and Alternative Dunn Index (ADI) [9]. It is already established that current high throughput technology has a significant impact on genomic and post-genomic studies including gene

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 680–687, 2013.

identification, disease diagnosis, drug discovery and toxicological research. For instance, the accurate identification of genes is essential for a successful diagnosis and treatment of a disease like lung cancer. One of the major challenges associated with lung cancer is the identification of disease mediating genes.

Incorporation of fuzzy set theory enables one to deal with uncertainties in different tasks of designing an intelligent system, arising from deficiency (e.g., vagueness, incompleteness) in information, in an efficient manner [13–15]. Fuzzy set theory has been applied to formulate several cluster validity indices like Partition Coefficient Index (PCI) (of the data, one may refer [1, 9]), Classification Entropy Index (CEI) [1], Partition Index (SCI) [11], Separation Index (SI) [11], Xie and Beni's Index (XBI) [16], Fukuyama and Sugeno Index (FSI) [17], Fuzzy Hypervolume Index (FHVI) [18], Dave's modification of the PC index (MPCI) [19], Partition Coefficient and Exponential Separation Index (PCAESI) [12], Index Based on Akaikes information criterion (AICI) [20], Compose Within and Between scattering Index (CWBI) [10], PBMF-Index (PBMFI) [21]. However, there is no instance of using cluster validity index, to our knowledge, which has been applied to the problem of finding disease mediating genes. The importance of the notion of fuzzy sets has been realized and successfully applied in almost all the branches of science and technology.

Thus the present paper proposed a novel cluster validity index, called Gaussian fuzzy index (GFI) in fuzzy set theoretic framework. The index involves measuring the average fuzzy intra-cluster distances over all the clusters, and inter-cluster distances between pairs of clusters. The smaller the value of GFI, better is the quality of the clusters. The effectiveness of GFI has been demonstrated on the human lung cancer in finding some possible genes mediating the disease [22]. Moreover, we have demonstrated superior capability of GFI, in identifying genes mediating the lung cancer, through an extensive comparative study of GFI with 8 existing validity indices namely C-index (CI) [6], Partition Coefficient Index (PCI) [1, 9], Partition Index (SCI) [11], Separation Index (SI) [11], Fukuyama and Sugeno Index (FSI) [17], Alternative Dunn Index (ADI) [9], Partition Coefficient and Exponential Separation Index (PCAESI) [12], Index Based on Akaikes information criterion (AICI) [20] on two clustering algorithms, viz., k-means [23] and fuzzy c-means (FCM) [24] with Euclidean distance as similarity measure. The results are appropriately validated using biochemical pathways. Moreover, we have considered simple pattern recognition problems to visualize the effectiveness of GFI. Thus the comparative performance of the cluster validity indices to identify good and meaningful clusters, has been evaluated internally through identification of disease (lung cancer) mediating genes. The external evaluation of the indices has been made through consulting pathway database and it has been shown that both the forms of evaluation are comparable.

2 Methodology

Let us consider a set of samples $U = \{\mathbf{x}_k | k = 1, 2, \ldots, n\}$ that are distributed in l clusters $C_1, C_2, \ldots, C_l$. These clusters have been obtained by a clustering algorithm.

2.1 Gaussian Fuzzy Index (GFI) for Cluster Validation

We now define a cluster validity index, called Gaussian Fuzzy Index, that will demonstrate the goodness of the results obtained by a clustering algorithm. Gaussian Fuzzy Index (GFI) is defined as

$$GFI = \frac{E'}{1+E} \tag{1}$$

where E' is given by

$$E' = \frac{2}{l(l-1)} \sum_{\substack{k,j=1 \\ k \neq j}}^{l} \mu_k(\mathbf{c}_j) \tag{2}$$

and E defined by

$$E = \frac{1}{l} \sum_{k=1}^{l} \frac{1}{|C_k|} \sum_{\mathbf{x}_p \in C_k} \mu_k(\mathbf{x}_p) \tag{3}$$

The term $\mu_k(\mathbf{c}_j)$ represents the membership value indicating the degree of belongingness of the center of jth cluster C_j to kth cluster C_k, and l stands for the number of resulting clusters. The membership function we have considered here is of Gaussian type, and is defined as

$$\mu_k(\mathbf{c}_j) = \exp(-\frac{||\mathbf{c}_j - \mathbf{c}_k||^2}{L^2}) \tag{4}$$

Here $\mathbf{c}_k$ and $\mathbf{c}_j$ are the kth and jth cluster centers respectively. The term L indicates the maximum distance between two objects in the set U (i.e., set of all the data objects). Thus L is represented by

$$L = \max_{\substack{\mathbf{x}_p, \mathbf{x}_{p'} \in U \\ p \neq p'}} ||\mathbf{x}_p - \mathbf{x}_{p'}|| \tag{5}$$

It is to be mentioned here that the elements are chosen from normed linear space. Similarly, $\mu_k(\mathbf{x}_p)$, the membership value of pth sample $\mathbf{x}_p$ to kth cluster C_k, is defined as

$$\begin{aligned} \mu_k(\mathbf{x}_p) &= \exp(\frac{-||\mathbf{x}_p - \mathbf{c}_k||^2}{\sigma_k^2}), \text{ where } \mathbf{x}_p \in C_k \\ &= 0, \qquad\qquad\qquad\quad \text{otherwise} \end{aligned} \tag{6}$$

The term σ_k is the diameter of kth cluster C_k, and is defined as

$$\sigma_k = \max_{\mathbf{x}_p, \mathbf{x}_{p'} \in C_k} ||\mathbf{x}_p - \mathbf{x}_{p'}|| \tag{7}$$

We say that a set of clusters to be good if the inter-cluster distances are large and intra-cluster distances are small. Here, E (in Equation (3)) represents the average fuzzy intra-cluster distance over all the clusters. The value of E lies in [0,1]. $E = 0$ represents the highest average fuzzy intra-cluster distance over all the clusters. It is to be mentioned that since E can be zero, we have added 1 in the denominator of Equation (1). On the other hand, the lowest average fuzzy

intra-cluster distance over all the clusters is obtained at $E = 1$. Likewise, E' (in Equation (2)) represents the average fuzzy distance among the cluster centers or average fuzzy inter-cluster distance. As in the case of E, E' lies in [0,1]. $E' = 0$ indicates the highest fuzzy inter-cluster distance over all the pairs of clusters. On the other hand, the lowest average fuzzy inter-cluster distance over all the pairs of clusters corresponds to $E' = 1$. Thus, a set of clusters is said to be good if the value of GFI is minimum. In other words, lower the value of GFI, better is the set of clusters obtained by an algorithm.

2.2 Comparative Study of Cluster Validity Indices and Selection of Possible Disease Mediating Genes

The performance of GFI is compared with 8 cluster validity indices. For this comparative study, we consider the following work flow.

- *Step I*: **Generation of clusters** – A clustering algorithm $\mathcal{C}$ is applied on a gene expression data with the different number (k for k-means and c for fuzzy c-means) of clusters as its input. Here we have considered these numbers ranging from 2 to 20. It is to be noted that the gene expression profiles for normal and diseased states are considered separately, and the number of clusters to be generated in the diseased state is kept equal to that for normal state.
- *Step II*: **Selection of the best k-value (or c-value) using a cluster validity index** – Among these 8 k-values (or c-values), the best k-value (or c-value) has been selected based on a cluster validity index.
- *Step III*: **For each k-value (or c-value) and for the clustering algorithm $\mathcal{C}$, the following steps are performed.** It is to be mentioned here that we have considered $k = 2, 3, \ldots 20$, in Step I, for each clustering algorithm. In this step (Step III), we consider the same k-values as in Step I.
  - *Step III.1*: **Determining corresponding clusters** – Clusters obtained in Step I using the clustering algorithm $\mathcal{C}$ for a k-value (or c-value) for both normal and diseased states need to be matched. Let C_i^N and C_j^D be ith and jth clusters, obtained by the clustering algorithm $\mathcal{C}$ for a k-value (or c-value), for normal and diseased states respectively. We say that the cluster C_i^N, for normal state, corresponds to cluster C_j^D, for diseased state, if $|(C_i^N \cap C_j^D)|$ is maximum over $j = 1, 2, \ldots, j, \ldots, k$.
  - *Step III.2*: **Identifying altered gene clusters** – We call a gene to be an altered gene if the gene is in C_i^N and C_j^D, where $i \neq j$. Thus, we can write an altered gene set $A_i = \cup_{j=1\ j\neq i}^{k} (C_i^N \cap C_j^D)$, for C_i^N. Thus, altered gene sets or altered clusters (*i.e.*, $A_1, A_2, \ldots A_{k-1}, A_k$) are generated from k normal clusters.
  - *Step III.3*: **Scoring an altered gene set** – Let the number of matched genes in altered gene sets $A_1, A_2, \ldots A_{k-1}, A_k$ be $l_1, l_2, \ldots, l_{k-1}, l_k$, respectively. Thus, the score for S_k is defined as

$$S_k = \tfrac{1}{k} \times \sum_{i=1}^{k} \frac{l_i}{|A_i|} \times 100\% \tag{8}$$

Higher the value of S_k, better is the matching. In other words, if S_k, for a clustering algorithm and cluster validity index, is high, the index is highly capable of identifying genes mediating the lung cancer provided the said clustering algorithm is used.

– *Step IV*: **Determining the best k-value (or c-value) and selection of some possible genes mediating lung cancer**– Let the k-value (or c-value) for which S_k score are maximum be K_S. Let the best k-value (or c-value) obtained by a cluster validity index I be K_I. A cluster validity index performs the best if and only if $|K_S - K_I| = 0$.

Hence, we say that a cluster validity index I_1 is better than I_2 if

$$|K_S - K_{I_1}| < |K_S - K_{I_2}| \tag{9}$$

3 Results

The effectiveness of GFI has been demonstrated on human lung cancer gene expression dataset. The performance of GFI has been compared extensively with 8 indices. Moreover, we have provided results on two pattern recognition problems to ease of depiction of the effectiveness of GFI.

3.1 Description of the Dataset

Human lung gene expression data is obtained by oligonucleotide microarray experiments for Ann Arbor tumors and normal lung samples [22]. In this data set, there are 7129 genes (more specifically, Affymetrix probe-sets) for 86 lung tumor and 10 normal lung samples. The gene expression profiles represent 86 primary lung adenocarcinoma, including 67 stage I and 19 stage III tumors, as well as 10 neoplastic lung samples.

3.2 Comparative Results Using Pathway Database

A biological pathway corresponds to a specific function or a group of functions. Such a pathway is actually a cascade of reactions involving proteins and metabolites. The proteins are again end product of genes. For example, if a set of proteins (genes) are involved in a pathway and the pathway is responsible for apoptosis (the process by which cells die) related function, then malfunctioning of the pathway may lead to inhibit apoptosis. Thus there may be an overgrowth of cells leading to lung cancer. In NCBI, we have got pathway related information from bio-system database[1]. Here we have found some lung cancer specific pathways like non-small cell lung cancer, small cell lung cancer. We have identified the genes (proteins) involved in these pathways. If the genes (i.e., corresponding proteins) in the altered gene sets are involved in such a pathway, we say that GFI

[1] `http://www.ncbi.nlm.nih.gov/Database`

has correctly identified some possible genes mediating the lung cancer. Higher the number of such match, better is cluster validity index.

For lung expression data, we have considered $k = 2$ to $k = 20$ for k-means and $c = 2$ to $c = 20$ for fuzzy c-means clustering algorithms. Using k-means algorithm, we have found that the best results of GFI corresponds to $k = 10$. This result produces $|K_S - K_{GFI}| = 0$ for GFI. On the other hand, applying k-means algorithm, the best results have been obtained for $k = 8$ by CI; $k = 11$ by SI; $k = 12$ by SCI and FSI; $k = 13$ by ADI; $k = 14$ by PCI; $k = 15$ by PCAESI and AICI. We have also got maximum scores of S (Equation (8)) for $k = 10$ using k-means ($S_{10} = 91.74\%$). Figure 1 depicts that GFI performs the best with respect to 8 cluster validity indices for k-means algorithm on lung expression dataset. It is to be noted that the other validity indices have generated their best values between $k = 8$ and $k = 12$. Hence, we can say that for the lung expression data considered here, the high quality clusters have been generated by k-means algorithm between $k = 8$ and $k = 12$.

Similarly, applying fuzzy c-means, the best result generated by GFI is $c = 13$, such that $|K_S - K_{GFI}| = 1$. On the other hand, the best results for $c = 10$ by PCI; $c = 11$ by ADI; $c = 12$ by CI and AICI; $c = 13$ by SCI; $c = 15$ by PCAESI and FSI; $c = 16$ by SI. The maximum score ($S_{14} = 93.13\%$) has been generated for $c = 14$. From Figure 1, it is clearly seen that our proposed validity index (GFI) generates the best result for $c = 13$, which is very close to the result generated by the above indices. It is to be noted that, for fuzzy c-means algorithm, all the 8 validity indices have shown their best results between $c = 12$ and $c = 15$.

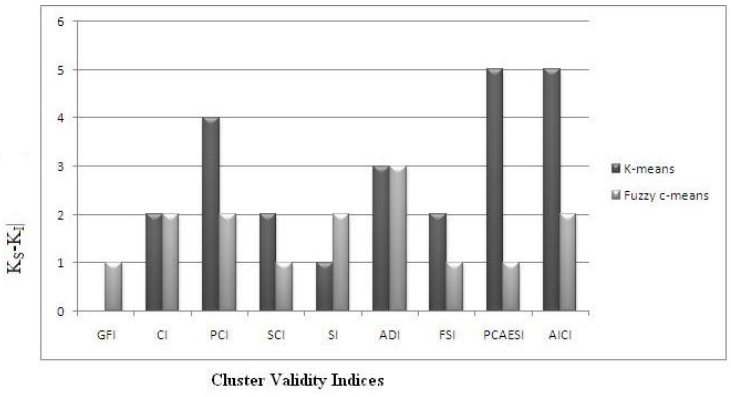


Fig. 1. Comparative values of $|K_S - K_I|$ for different cluster validity indices using pathway database for k-means and fuzzy c-means clustering algorithms on lung cancer dataset

3.3 Selection of Some Possible Genes Mediating Lung Cancer

Here we report the genes in the altered gene sets whose expression values have deviated from normal to disease states of human lung cancer dataset. From human lung expression data, the proposed index (GFI) has identified the genes (from altered gene set) like EGFR, TNF, TNFSF11, RIMS2, KRAS, HLA-G, TP53, VEGFA, IL6, CDKN2A, STAT3, CDH1, TGFB1, IL10, IL8, PTEN, MYC, IGFBP3, TNFSF10, CASP3, CD44, IGF1R. Moreover, we can say that the aforesaid altered genes have a significant role in the development of the aforesaid types of lung adenocarcinoma. In other words, we can say the aforesaid genes may have a strong influence in mediating the lung cancer. It is interesting to note that the proposed index GFI has been able to identify more responsible genes (for mediating the lung cancer) supported by a wide range of earlier investigations.

4 Conclusions

In this article, we have developed a cluster validity index, called Gaussian Fuzzy Index (GFI), using the notion of fuzzy set theory. The proposed index involves the average fuzzy intra-cluster distances over all the clusters, and inter-cluster distances between pairs of clusters. The index GFI is formulated in such a way that its minimization leads to minimization of fuzzy intra cluster distance and maximization of fuzzy inter cluster distance. The smaller the value of GFI signifies the better quality of the clusters.

The effectiveness of the index GFI has been demonstrated using two clustering algorithms on human lung cancer dataset [22]. Moreover, we have made an extensive experiment for identifying the important genes from a gene expression data. This concept leads to predict some possible disease mediating genes for certain human lung cancer. The results have appropriately been validated using biochemical pathways. We have also implemented different cluster validity indices to demonstrate superior capability of the GFI over the others.

References

1. Bezdek, J.C.: On clustering validation techniques. J. Cybernet. 17, 58–73 (1974)
2. Deborah, L.J., Baskaran, R., Kannan, A.: A survey on internal validity measure for cluster validation. IJCSES 1, 85–102 (2010)
3. Dunn, J.C.: Well separated clusters and optimal fuzzy partitions. J. Cybern. 4, 95–104 (1974)
4. Davies, D.L., Bouldin, D.W.: A cluster separation measure. IEEE Trans. Pattern Anal. Machine Intell. 1, 224–227 (1979)
5. Rousseeuw, P.J.: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65 (1987)
6. Hubert, L., Schultz, J.: Quadratic assignment as a general data-analysis strategy. British Journal of Mathematical and Statistical Psychologie 29, 190–241 (1976)

7. Goodman, L., Kruskal, W.: Measures of associations for cross-validations. J. Am. Stat. Assoc. 49, 732–764 (1954)
8. Pauwels, E.J., Frederix, G.: Finding salient regions in images: nonparametric clustering for image segmentation and grouping. Computer Vision and Image Understanding 75, 73–85 (1999)
9. Trauwaert, E.: On the meaning of dunn's partition coefficient for fuzzy clusters. Fuzzy Sets Systems 25, 217–242 (1988)
10. Yun, X.U., Brereton, G.R.: A comparative study of cluster validation indices applied to genotyping data. Chemometrics and Intelligent Laboratory Systems 78, 30–40 (2005)
11. Bensaid, A.M., Hall, L.O., Bezdek, J., Clarke, L.P., Silbiger, M.L., Arrington, J.A., Murtagh, R.F.: Validity-guided (re) clustering with applications to imige segmentation. IEEE Transactions on Fuzzy Systems 4, 112–123 (1996)
12. Wu, K., Yang, M.: A cluster validity index for fuzzy clustering. Pattern Recognition Lett. 26, 1275–1291 (2005)
13. Zadeh, L.A.: Fuzzy sets. Information and Control 8, 338–353 (1965)
14. Zadeh, L.A.: A fuzzy-set-theoretic interpretation of linguistic hedges. Journal of Cybernetics 2, 4–34 (1972)
15. Bandler, W., Kohout, L.J.: Fuzzy power sets and fuzzy implication operators. Fuzzy Sets and Systems 4, 13–30 (1980)
16. Xie, X.L., Beni, G.A.: Validity measure for fuzzy clustering. IEEE Trans. PAMI 3, 841–846 (1991)
17. Fukuyama, Y., Sugeno, M.: A new method of choosing the number of clusters for the fuzzy c-means method. In: Proceeding of fifth Fuzzy Syst. Symp., pp. 247–250 (1989)
18. Gath, I., Geva, A.B.: Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Machine Intell. 11, 773–781 (1989)
19. Dave, R.N.: Validating fuzzy partition obtained through c-shells clustering. Pattern Recognition Lett. 17, 613–623 (1996)
20. Akaike, H.: A bayesian extension of the minimum aic procedure of autoregressive model fitting. Biometrika 66, 237–242 (1979)
21. Pakhira, M., Bandyopadhyay, S., Maulik, U.: A study of some fuzzy cluster validity indices, genetic clustering and application to pixel classification. Fuzzy Sets and Systems 155, 191–214 (2005)
22. Beer, G.D., et al.: Gene-expression profilespredict survival of patients with lung adenocarcinoma. Nature Medicine 8, 816–823 (2002)
23. Dubes, R.C., Jain, A.K.: Algorithms for clustering data. Prentice Hall (1988)
24. Bezdek, J.: Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York (1981)
25. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugenics 7, 179–188 (1936)
26. Gibbons, F.D., Roth, F.P.: Judging the quality of gene expression-based clustering methods using gene annotation. Genome Research 12, 1574–1581 (2002)

Semi-supervised Self-organizing Feature Map for Gene Expression Data Classification

Moumita Roy, Anwesha Law, and Susmita Ghosh

Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

Abstract. In this article, a one-dimensional self-organizing feature map (SOFM) neural network integrated with semi-supervised learning is used to predict the class label of gene expression data under the scarcity of the labeled patterns. Iterative learning of the semi-supervised SOFM network is carried out using a few labeled patterns along with some selected unlabeled patterns. The unlabeled patterns, for which the maximum target value is greater than a threshold, are selected as the confident ones. Results are found to be encouraging.

Keywords: Semi-supervised learning, gene expression data, self-organizing feature map.

1 Introduction

Microarray technology is an important tool used to monitor expression levels of genes of a given organism. A microarray is a glass slide onto which DNA molecules are fixed at specific locations called spots or features. A microarray may contain thousands of spots and each spot may contain a few million copies of identical DNA molecules that uniquely correspond to a particular gene. Gene expression data is mainly used to diagnose (predict class labels) a patient depending on his medical condition [1]. One of its major applications is precise and early diagnosis of cancer malignancies [5] which is quite difficult but very crucial for successful treatment. Gene expression technology integrated with some accurate statistical methods could play a better role for cancer diagnosis than traditional clinical factors. Given microarray experiments and some information about the outcome of the disease from some former patients, the task of class prediction (malignancies present or not) is to learn the relation between the gene expression levels and their corresponding outcome. Afterwards, a diagnosis of disease development for new patients can be carried out using their gene expression profiles. There are mainly two approaches used for class prediction of gene expression data: supervised and unsupervised. Between these, applicability of supervised method [5] is lesser due to insufficient available knowledge; whereas, prior information is not necessary for unsupervised methods and hence they are widely applied to find groups of co-regulated genes, categorize them, and predict information about an independent sample [11]. A number of studies have been done in this aspect [3,10].

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 688–694, 2013.

Microarray experiments generate large datasets with expression values for thousands of genes (features), but usually for a very few samples (i.e., patients). The labeling of the samples require expert knowledge, which, in turn, gives rise to a situation of (labeled) data scarcity. Due to insufficiency of training samples, supervised methods cannot be developed. Also, if an unsupervised method is used, knowledge of labeled samples, though very little, will remain unused. In this situation, semi-supervised approach [4,6] could be opted instead of an unsupervised or a supervised one. Semi-supervision uses a small amount of labeled patterns with abundant unlabeled ones for learning, and integrates the merits of both supervised and unsupervised strategies to make full utilization of collected patterns [4]. Usefulness of semi-supervised approaches is already explored to predict the class label of gene expression data [9].

Artificial neural networks (ANNs) [7], one of the powerful soft-computing tools, have widespread applications in pattern recognition and image processing [2]. ANNs have the most appreciable quality of learning with or without supervision. Neural networks use nonparametric statistical approaches for pattern classification where no prior knowledge for input pattern distribution is required. Due to massive parallel nature of neural networks, results can be obtained in less time. Applications of various types of neural networks, e.g., self-organizing feature maps (SOFM) [8], multilayer perceptron [7] are already made in the area of class prediction [10,12] of gene expression data. But, as to the knowledge of the authors, neural network based semi-supervised approaches are not found for gene expression data classification. This motivated us to pursue the present study using neural networks in semi-supervised learning platform to improve the performance of class prediction of gene expression data.

In the proposed work, SOFM neural network [6,7] embedded with semi-supervised learning (termed as, SS-SOFM) is used for class prediction of microarray data using a few labeled data along with the selected (confident) unlabeled patterns. To check the effectiveness of the proposed approach, experiments are carried out on five preprocessed gene expression datasets [5]. Comparative analysis has been carried out between the proposed semi-supervised strategy and the corresponding unsupervised method using SOFM.

2 Kohonen's Self-organizing Feature Map

Self-organizing feature map (SOFM) network [7,8] learns iteratively and generates topological map of input patterns gradually. The network architecture consists of two layers: input and output. The number of neurons in the input layer is equal to the number of features of the input pattern. Here, one dimensional output layer is used and the number of neurons in the output layer is equal to the number of classes. Each neuron in the output layer is connected to all the neurons in the input layer. The connection weights are initialized randomly between $[0, 1]$. Training of the SOFM is continued epoch by epoch till convergence by following three consecutive steps: competition, co-operation and weight updating.

Let, the p^{th} input pattern be denoted by $\overrightarrow{X_p} = [x_{p,1}, x_{p,2}, ..., x_{p,y}]$, where y is the number of features of the input pattern. The y-dimensional weight vector of the i^{th} neuron in the output layer connected with all the neurons of the input layer at epoch ep is represented by $\overrightarrow{W_i^{ep}} = [w_{i,1}^{ep}, w_{i,2}^{ep}, ..., w_{i,y}^{ep}]$ $(i = 1, 2, ..., z$, where z is the total number of output neurons).

In the first step, the winner neuron j at epoch ep in the output layer is selected as the best-matched neuron for the p^{th} input pattern where the similarity measure $(\overrightarrow{W_j^{ep}} \cdot \overrightarrow{X_p})$ is the maximum. Now, the winning neuron co-operates with its neighbors. Let $h_{i,j}^{ep}$ denote the influence of the winning neuron j on its topological neighbor i at epoch ep and $D_{i,j}$ denotes the lateral distance between the winning neuron j and its neighbor i. The topological neighborhood $h_{i,j}^{ep}$ is defined in a way such that it attains the maximum value at the winning neuron j for which $D_{i,j}$ is zero and decreases monotonically with increasing lateral distance $D_{i,j}$. Here, $h_{i,j}^{ep}$ is defined as follows:

$$h_{i,j}^{ep} = \exp\left(-\frac{{D_{i,j}}^2}{2(\sigma^{ep})^2}\right), \tag{1}$$

where σ^{ep} represents the radius of the topological neighborhood at epoch ep. The size of this radius decreases with increase in ep.

The synaptic weight is then updated using the following equation:

$$\overrightarrow{W}_i^{ep+1} = \overrightarrow{W}_i^{ep} + h_{i,j}^{ep}\ \eta^{ep}\ (\overrightarrow{X_p} - \overrightarrow{W}_i^{ep}), \tag{2}$$

where, η^{ep}, the learning rate in epoch ep, decreases with increase in ep.

Table 1. Algorithmic representation of the proposed work

Step 1:	Collect a few labeled patterns from each of the classes.
Step 2:	Initialize (randomly) the connection weights of SOFM.
Step 3:	Train the SOFM with the few collected labeled patterns until convergence.
Step 4:	For all the unlabeled patterns, predict the output for each of the classes by passing them through the (trained) SOFM.
Step 5:	For all the unlabeled patterns, estimate the target values using the output values of itself and its K nearest neighbours.
Step 6:	Select the confident unlabeled patterns, for which the maximum target value is greater than a threshold.
Step 7:	Train the SOFM with the same set of labeled patterns along with the set of selected confident unlabeled patterns.
Step 8:	Repeat Steps 4-7 until convergence. At convergence, goto Step 9.
Step 9:	Assign hard class labels to all the unlabeled patterns.

3 The Proposed Semi-supervised SOFM Algorithm

As mentioned in Section 1, in the present work semi-supervised SOFM neural network is used for class prediction of gene expression data where a few labeled

patterns are available and they are required for semi-supervised learning. For the labeled patterns, if it belongs to the l^{th} class, then the target values and output values for the said class are both assigned to 1 and for the rest of the classes these values are assigned to 0 and are kept fixed during training as well as during testing phases. The algorithmic representation of the proposed methodology is given in Table 1 and detailed description is presented in subsequent sections.

3.1 Training of the SOFM

In the proposed work, learning process of SOFM is modified to incorporate the label information of the patterns. During training, the patterns are fed to the network one by one. At epoch ep (in the training step st), for p^{th} pattern, the output value $s_{p,l}^{ep}$ in the l^{th} class (here, each of the classes is represented by a neuron in the output layer) is estimated by computing the similarity measure between $\overrightarrow{X_p}$ and $\overrightarrow{W_l^{ep}}$, and is taken as,

$$s_{p,l}^{ep} = \overrightarrow{X_p} \cdot \overrightarrow{W}_l^{ep} = \sum_{k=1}^{y} x_{p,k} \cdot w_{l,k}^{ep}. \tag{3}$$

As already mentioned in Section 1, in the present work, iterative learning of SOFM is carried out using the labeled as well as the selected unlabeled patterns. Unlike the traditional SOFM, here, an output neuron is treated as the winning neuron for a given input pattern $\overrightarrow{X_p}$ if the target value of the pattern in the corresponding class is the maximum. In this way, we have incorporated the label information during learning of the network. Thereafter, the weight vectors of the winning neuron as well as its neighboring neurons are updated using equation (2). Learning using labeled patterns is continued epoch by epoch until convergence. To check convergence, the sum of square error between the target value and the estimated output value at epoch ep, denoted as O^{ep}, is calculated using

$$O^{ep} = \sum_{p=0}^{N} \sum_{l=1}^{C} (t_{p,l}^{st} - s_{p,l}^{ep})^2, \tag{4}$$

where N is the total number of input patterns and $t_{p,l}^{st}$ is the target value of the p^{th} pattern for the l^{th} class at training step st. Here, C denotes the number of classes. The weight updating is performed until $(O^{ep} - O^{ep-1}) < \delta$, where δ is a very small positive quantity. After each updating, the components of the weight vector $\overrightarrow{W_l^{ep}}$ are normalized to lie in $[0, 1]$. After every epoch, the values of learning rate and the size of the topological neighborhood are decreased.
Since, in the first training step, we do not have defined target values for the unlabeled patterns, the connection weights of the SOFM network are updated using the labeled patterns only.

3.2 Estimation of Target Value for Unlabeled Patterns

After convergence, the unlabeled patterns are presented to the network and the output in all the classes for each of the unlabeled patterns are predicted using

equation (3). Then, soft class label for each unlabeled pattern is estimated by averaging the corresponding output values of its K nearest neighbors thereby incorporating the labeled information from neighbors. The estimated soft class label is used for assigning the target value of the unlabeled patterns.

3.3 Collection of Unlabeled Patterns and Iterative Training of the SOFM

As mentioned earlier, in the present work, learning of SOFM is carried out iteratively using the labeled patterns along with the selected unlabeled patterns. Here, an unlabeled pattern, for which its maximum target value is greater than a threshold, is selected as the confident one. The threshold is obtained by averaging the summation of maximum target values for all the unlabeled patterns.
Training of SOFM, estimation of the target values for all the unlabeled patterns, and collection of the (confident) unlabeled patterns for the next training step are carried out iteratively until convergence or the number of training steps exceeds a prespecified value. To check convergence, at the end of each training step, the sum of square error (ξ_{st}) between the predicted support values using the (trained) SOFM network and the updated target values using the estimated support values of K nearest neighbors for all the unlabeled patterns is computed. The training steps are performed until $(\xi_{st} - \xi_{st-1}) < \epsilon$, where ϵ is a very small positive quantity. After convergence, the hard class labels are assigned to the unlabeled patterns depending on their target values.

4 Experimental Details and Analysis of Results

4.1 Datasets Used

To evaluate the effectiveness of the proposed SOFM based semi-supervised method (denoted as, SS-SOFM), experiments are conducted on five gene expression datasets denoting different types of cancers. Preprocessed versions of the datasets are available in [5]. In the present work, an input pattern consists of gene expression levels of a particular patient. The feature values of the input patterns are normalized between 0 to 1. Detailed description of the data sets are given below:

A. Brain tumor dataset: This dataset contains microarray gene expression profiles of 42 patients having 5 different types of tumors of the central nervous system. It has 10 medulloblastomas, 10 malignant gliomas, 10 atypical teratoid/rhabdoid tumors (AT/RTs), 8 primitive neuro-ectodermal tumors (PNETs) and 4 human cerebella. The raw data consists of 5,597 genes.

B. Colon cancer dataset: This dataset consists of expression levels of 40 tumor and 22 normal colon tissues for 2000 human genes.

C. Leukemia dataset: Gene expression levels of 72 patients either suffering from acute lymphoblastic leukemia (ALL, 47 cases) or acute myeloid leukemia (AML, 25 cases) are present in this dataset. It has a total of 3,571 genes.

D. Lymphoma dataset: There are a total of 62 samples, and the expression of 4,026 genes. This dataset contains gene-expression levels of the 3 most prevalent adult lymphoid malignancies, i.e., 42 samples of diffuse large B-cell lymphoma (DLBCL), 9 observations of follicular lymphoma (FL), and 11 cases of chronic lymphocytic leukemia (CLL).

E. Prostate cancer dataset: It has expression of 6,033 genes each, from 102 samples, of which 52 are of prostate tumors and 50 are from non-tumor prostate samples.

4.2 Results and Analysis

To assess the effectiveness of the proposed approach, experiments are conducted on five gene expression datasets. Performance of the proposed technique is compared with that of the corresponding unsupervised one using one dimensional SOFM neural network. To execute the proposed semi-supervised algorithm, for all the data sets, a single pattern from each of the classes are chosen randomly as labeled patterns. To find out K number of nearest neighbors, the value of K has been set to 9 for all the datasets. Results corresponding to semi-supervised and unsupervised techniques are depicted in Table 2. The best percentage of accuracy value (denoted as, PA) obtained over 20 different simulations and its corresponding Rand and Jacard indices are used as the performance measuring criteria. The values of Rand and Jacard coefficients lie between [0,1] and for both the indices, higher value signifies better class prediction. From the results, it is seen that for all the cases the proposed methodology outperforms the corresponding unsupervised technique in terms of all the measuring indices used in our experiment. The results also strengthen the fact that the use of a very small number of labeled patterns increases the performance of the system significantly.

Table 2. Comparative results

Datasets	Techniques	#Training Patterns	PA	Rand	Jacard
Brain Tumor	Unsupervised	-	71.4286	0.8200	0.3673
	SS-SOFM	5	80.9524	0.8769	0.5810
Colon Cancer	Unsupervised	-	58.0645	0.5050	0.3514
	SS-SOFM	2	70.9677	0.5447	0.5348
Leukemia	Unsupervised	-	76.3889	0.6342	0.4771
	SS-SOFM	2	97.2222	0.9452	0.9051
Lymphoma	Unsupervised	-	70.9677	0.6790	0.4467
	SS-SOFM	3	88.7097	0.8297	0.6840
Prostate	Unsupervised	-	60.7843	0.4952	0.3352
	SS-SOFM	2	75.4902	0.6263	0.4548

5 Conclusion

In this paper, a semi-supervised self-organizing feature map neural network is designed for more accurate prediction of class information for gene expression data. It is achieved by improvising semi-supervision along with the self-organizing capability of Kohonen's network. The network is trained by using a small number of labeled patterns along with a large amount of unlabeled patterns. From the results, it has been found that the proposed algorithm has an edge over the corresponding unsupervised version. Like other semi-supervised methods, the proposed technique also computation intensive.

References

1. Bhattacharya, A., De, R.K.: Average correlation clustering algorithm (ACCA) for grouping of co-regulated genes with similar pattern of variation in their expression values. Journal of Biomedical Informatics 43(4), 560–568 (2010)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
3. Chandrasekhar, T., Thangavel, K., Elayaraja, E.: Gene expression data clustering using unsupervised methods. In: Proceedings of the 3rd International Conference on Advanced Computing (ICoAC), pp. 146–150 (2011)
4. Chapelle, O., Schölkopf, B., Zien, A.: Semi-supervised Learning. MIT Press (2006)
5. Dettling, M.: BagBoosting for tumor classification with gene expression data. Bioinformatics 20(18), 3583–3593 (2004)
6. Ghosh, S., Roy, M.: Modified self-organizing feature map neural network with semi-supervision for change detection in remotely sensed images. In: Kuznetsov, S.O., Mandal, D.P., Kundu, M.K., Pal, S.K. (eds.) PReMI 2011. LNCS, vol. 6744, pp. 98–103. Springer, Heidelberg (2011)
7. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice-Hall of India, New Delhi (2008)
8. Kohonen, T.: Self-Organizing Maps, 2nd edn. Springer, Berlin (1997)
9. Maraziotis, I.A.: A semi-supervised fuzzy clustering algorithm applied to gene expression data. Pattern Recognition 45(1), 637–648 (2012)
10. Nikkila, J., Törönen, P., Kaski, S., Venna, J., Castrén, E., Wong, G.: Analysis and visualization of gene expression data using Self-Organizing Maps. Neural Networks 15(8-9), 953–966 (2002)
11. Srinivas, V.R.: Bioinformatics: A Modern Approach, 2nd edn. Prentice-Hall of India Pvt. Ltd., New Delhi (2007)
12. Valentini, G.: Supervised gene expression data analysis using support vector machines and multi-layer perceptrons. In: Proceedings of the 6th International Conference on Knowledge-Based Intelligent Information & Engineering Systems (2002)

FlowSummary: Summarizing Network Flows for Communication Periodicity Detection

Neminath Hubballi and Deepanshu Goyal

Department of Computer Science & Engineering, Indian Institute of Technology Guwahati, Assam 781039, India
{neminath,g.deepanshu}@iitg.ernet.in

Abstract. Data summarization is an important technique to understand large datasets and discover useful patterns. In this paper we formulate a problem of summarizing network flow data to discover periodic communication behavior. An efficient implementation method for discovering periodic patterns is described in this paper and it has successfully discovered such patterns in a simulated and real application.

1 Introduction

Data summarization is an important technique for analyzing large dataset. It represents a large set of data points in a compressed format. There are many domain specific data summarization techniques proposed in the literature [10,9]. Similarly there are summarization techniques proposed for cyber attack detection [3] in the literature. There are limitations to existing techniques such as information loss and have limited scope of detecting a specific subset of cyber attacks. However in this paper we propose a method which is lossless and can be useful in detecting a range of cyber attacks. The summarization technique detects periodic communication behaviors of a host. It is shown that a large set of applications (both good and bad) exhibit periodic communication behaviors [2]. These are mostly automated communications and not driven by user actions. For example, an application may be polling the server for possible new updates. A backup database may be synchronizing the data with a main server at periodic intervals. On the other hand these automated communications may be due to a malicious application running in the host exporting sensitive data. From a system administrator's perspective identifying such communications is important as it will have implications from security and privacy. The proposed summarization technique is useful for detecting network anomalies thus the works in this category are briefly reviewed. There are many anomaly detection techniques such as only packet header based [8,6] flow based [5,7,3] and protocol anomaly detection [4]. Header based anomaly detection techniques derive some statistics based on fields available in packet header. These statistics forms a feature vector and using a suitable data mining algorithm, abnormal behaviors can be detected. Flow based techniques on the other hand rate the connection between two communication end points. A suitable rating function is used to differentiate abnormal connections w.r.t normal connections. Protocol anomaly detection techniques detect abnormalities against the specified protocol behavior.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 695–700, 2013.

2 Periodicity Detection

The proposed technique is based on the fact that, periodic communications exhibit very low variance and standard deviation considering their inter time differences.

Table 1. Transactions of Flows

Flow	Source IP	Destination IP	Time
F_1^1	123.123.123.123	208.208.208.208	01:02:01,01-01-2013
F_2^1	123.123.123.123	208.208.208.208	01:12:01,01-01-2013
F_1^2	123.127.123.127	208.109.208.109	06:12:01,01-01-2013
F_3^1	123.123.123.123	208.208.208.208	01:22:01,01-01-2013
F_2^2	123.127.123.127	208.109.208.109	06:32:01,01-01-2013
F_3^2	123.127.123.127	208.109.208.109	06:52:01,01-01-2013

On the other hand communication flows of random nature exhibit very high variance. Thus we use the standard deviation of communications established between end points (identified by IP addresses) as a measure to decide whether they have periodic communication or random communication. Standard deviation is given by the Equation 1.

$$SD = \sqrt{\Big(\frac{1}{M-1}\sum_{i=1}^{M}(X_i - \mu)^2\Big)} \tag{1}$$

In Equation 1, for any random variable X, X_i is the value taken by that random variable in one observation and μ is the mean of all the values taken by that random variable.

Table 2. Transactions with $DiffTime$

Flow	Source IP	Destination IP	$DifTime$
F_1^1	123.123.123.123	208.208.208.208	NA
F_2^1	123.123.123.123	208.208.208.208	10
F_1^2	123.127.123.127	208.109.208.109	NA
F_3^1	123.123.123.123	208.208.208.208	10
F_2^2	123.127.123.127	208.109.208.109	20
F_3^2	123.127.123.127	208.109.208.109	20

We define the communication end points as two hosts identified by two distinct IP addresses which have exchanged at least one packet between them. We say a flow has started whenever there is an exchange of packets between two hosts and this communication is represented in the form of $SrcIP$-$DestiIP$ pairs indicating source and destination IP addresses of hosts involved. We maintain four parameters along with these pairs and subsequently show that these parameters are sufficient to describe about the nature of interaction between the two hosts. Let the host at which network traffic is collected be $\mathcal{H}$. This can be done using a suitable software like tcpdump [1]. Let P_1, P_2, ... ,P_N be a series of packets exchanged with $\mathcal{H}$ representing several communications over a period of time and $IP_1, \ldots IP_I$ ($1 \leq I \leq N$) be the total number of peer hosts involved in communication with $\mathcal{H}$. For each peer host identified by an IP address IP_K ($1 \leq K \leq I$) there may be one or more flows and let these flows be represented as $F_1^{IP_K}, F_2^{IP_K}, \ldots, F_M^{IP_K}$ ($M \leq N - I + 1$). Let $t_1^{IP_K}, t_2^{IP_K}, \ldots, t_M^{IP_K}$ be the corresponding timestamps of these flows (identified by the time-stamp of first packet in

the flow). A set of flows can be tabulated as in Table 1 and we refer this as transaction data.

A flow indicates the beginning of a new communication and our aim is to identify the nature of such interactions in this network traffic. To achieve this we derive a parameter called $DiffTime$ - which is a time-stamp difference between two successive flows i.e., the difference between $t_l^{IP_K}$ and $t_{l+1}^{IP_K}$ in the time series. With $DiffTime$ calculated for each flow, the transactions in Table 1 can be shown as in Table 2. In Equation 1 $DiffTime$ of individual flows is considered for calculating the standard deviation. Standard deviation calculation requires all the $DiffTime$ values for the entire period. Normally to detect periodic communications a hosts traffic need to be collected for a longer duration. Network traffic is bulky for storage and processing, thus an efficient technique is provided for standard deviation calculation. This technique is motivated by a similar approach for summarization [11]. To describe the nature of communication we summarize the flows between two hosts as shown in Equation 2. And call it as a $FlowSummarry$

$$SrcIP, DstIP, LS, SS, M, t_l \tag{2}$$

where
$SrcIP$ - Source IP address (sender of the first packet in a flow)
$DstIP$ - Destination IP address (recipient of the first packet in a flow)
LS - Linear sum of $DiffTime$ i.e., $\sum_{l=1}^{M}(t_{l+1}^{IP_K} - t_l^{IP_K})$
SS - Squared sum of $DiffTime$ i.e., $\sum_{l=1}^{M}(t_{l+1}^{IP_K} - t_l^{IP_K})^2$
M - Number of flows seen between the two hosts during the period
t_l - Timestamp of the last flow between two hosts
Using the information in $FlowSummarry$, it is possible to calculate standard deviation of inter time difference $DiffTime$. Mean μ is calculated as shown in Equation 3.

$$Mean = \mu = \frac{LS}{M} \tag{3}$$

Similarly the calculation of variance can be deduced to Equation 7 from Equation 1

$$SD = \sqrt{\Big(\frac{1}{M-1}\sum_{i=1}^{M}(X_i - \mu)^2\Big)} = \sqrt{\Big(\frac{1}{M-1}\Big[\sum_{i=1}^{M}(X_i^2 + \mu^2 - 2X_i\mu)\Big]\Big)} \tag{4}$$

$$SD = \sqrt{\Big(\frac{1}{M-1}\Big[\sum_{i=1}^{M}X_i^2 + \sum_{i=1}^{M}\mu^2 - 2\sum_{i=1}^{M}X_i\mu\Big]\Big)} \tag{5}$$

$$SD = \sqrt{\Big(\frac{1}{M-1}\Big[\sum_{i=1}^{M}X_i^2 + M*\mu^2 - 2\mu\sum_{i=1}^{M}X_i\Big]\Big)} \tag{6}$$

$$SD = \sqrt{\Big(\frac{1}{M-1}\{SS + M*\big(\frac{LS}{M}\big)^2 - 2*LS*\frac{LS}{M}\}\Big)} \tag{7}$$

With respect to a host $\mathcal{H}$, all the flows originating from the host will have $\mathcal{H}$'s IP address as source IP address and different destination addresses. On the other hand all the flows intended towards $\mathcal{H}$ will have different source addresses and $\mathcal{H}$'s IP address as destination address. This allows us to represent the interaction of $\mathcal{H}$, in the form of a graph.

3 Experimental Results

There are two types of experiments done, first is on a simualted application behavior and the other is with a real application exhibiting periodic behavior. For both the experiments network traffic was logged using tcpdump on a host connecting to ISP via a DSL broadband.

Table 3. Statistics for Artificial Dataset

Source IP	Destination IP	Flow Count	Std Devtn
192.XX.XX.4	202.XX.XX.6	4025	2.90
202.XX.XX.6	192.XX.XX.4	4024	2.86
192.XX.XX.4	216.XX.XX.148	12	0.88
184.XX.XX.29	192.XX.XX.4	12	7.02
192.XX.XX.4	184.XX.XX.29	13	6.97
174.XX.XX.114	192.XX.XX.4	14	3.97
192.XX.XX.4	174.XX.XX.114	14	3.99
173.XX.XX.51	192.XX.XX.3	14	5.05
192.XX.XX.4	74.XX.XX.54	14	5.74
74.XX.XX.54	192.XX.XX.4	14	5.85
192.XX.XX.4	115.XX.XX.38	15	1.54
202.XX.XX.246	192.XX.XX.4	15	1.83
192.XX.XX.4	119.XX.XX.10	16	0.25
119.XX.XX.10	192.XX.XX.4	16	0.25
209.XX.XX.191	192.XX.XX.4	18	0.24
74.XX.XX.162	192.XX.XX.4	20	4.56
192.XX.XX.4	74.XX.XX.162	20	4.58
174.XX.XX.139	192.XX.XX.4	21	1.41
115.XX.XX.38	192.XX.XX.4	21	1.53
174.XX.XX.139	192.XX.XX.4	21	1.74
192.XX.XX.4	50.XX.XX.165	26	6.49
50.XX.XX.165	192.XX.XX.4	26	7.50

In the first experiment, periodic communication behaviour was simulated with a Perl script generating requests for a web page for every 100 seconds. To make it close to real world traffic a small random number (between 1 to 10) was added to the base period so that there is a slight variation in the initiated communication. This expriment was run for 5 days along with other normal user usage. For the second experiment network traffic for real application having periodic behavior for 7 hours along with other normal browsing activity of end user was collected. A sport web page which shows live score card during a cricket match and also live text commentary was chosen as a real application. Normally such pages refresh the scorecard automatically using a html refresh directive (with a timer) thus the communication between web site and the local computer is periodic in nature and this exactly suit our experiments.

1)Artificially Generated Dataset: By analyzing simulated traffic we generated $FlowSummarry$ as described in previous section. For this experiment we assume a standard deviation of less than 10 to be a periodic communication. Table 3. shows list of periodic communications identified by $FlowSmmary$. First and second column represent the source and destination IP addresses of flows and third column shows the number of flows between these end hosts and last column shows the standard deviation. First two rows in Table 3. represent two genuine periodic communications with their standard deviations. The first entry shows communication from the host to the web server and second entry shows the connection in other direction. We can notice that, standard deviation for both the communications is around 2.9. with total number of flows being around 4000, in each direction.

Table 4. $FlowSummary$ Groups

Instances	Flow Count	SD Range	FP
337	001 - 002	NA	NA
220	003 - 005	000.47 - 10366.00	NA
178	006 - 010	000.30 - 08640.00	NA
089	011 - 015	002.32 - 05519.60	10
044	016 - 020	002.32 - 05212.53	05
030	021 - 025	001.41 - 04513.13	03
026	026 - 030	006.49 - 01044.70	02
012	031 - 035	005.72 - 02482.44	00
011	036 - 040	200.69 - 02415.39	00
008	041 - 045	011.17 - 00801.18	00
005	046 - 050	010.53 - 00085.00	00
013	051 - 100	031.90 - 01886.23	00
002	>100	002.86 - 00002.90	00
975	Total		

Table 4. shows the summary of communications during the period of monitoring. We observed a total of 975 distinct TCP flows i.e 975 distinct IP addresses being contacted by the host monitored (there are several flows for each of these IP addresses) in the traffic. This table shows the groups of flows based on the number of flows seen for a particular end host pair. Column 1 shows number of distinct end host pairs and column 2 shows the range for number of flows for a particular pair. For example row 1 is interpreted as follows. There are 337 distinct $SrcIP$-$DstIP$ pairs each having either one or two flows between them. Column 3 in this table shows the standard deviation range for the corresponding $FlowSummarry$ group. Column 4 shows the number of false positives in that group. These are the flows identified to be periodic by the algorithm and essentially the same flows shown in Table 3. except the first 2 entries. To qualify for periodic communication we use minimum threshold for number of flows to be at least 10. As we notice from the table nearly 60% of these $FlowSummarry$s are having a flow count of less than 5 (majority of them with count of 1). A closer observation, revealed that majority of these is communications originated from searches. This is because when the user is searching for particular information, within a short period of time it is likely that she searches for it repeatedly (till she finds useful information that she wanted). Typical user behavior is to open the search page and see if the information is what she wanted. Once the page is opened a connection is established with corresponding server.

Table 5. Statistics for Real Application Dataset

Source IP	Destination IP	Flow Count	Std Devitn
192.XX.XX.4	122.XX.XX.184	417	0.40
122.XX.XX.184	192.XX.XX.4	415	0.41
192.XX.XX.4	96.XX.XX.48	6	7.9
96.XX.XX.48	192.XX.XX.4	6	8.5
192.XX.XX.4	119.XX.XX.10	4	9.1
119.XX.XX.10	192.XX.XX.4	4	9.2
192.XX.XX.4	115.XX.XX.6	6	9.9
115.XX.XX.6	192.XX.XX.4	6	9.9

Out of 975 distinct communications a total of 20 false positives and there are 2 true positives. This is considered to be reasonably good performance even when we pick the threshold for standard deviation as high as 10 when the real periodic flows standard deviation is around 3. By reducing its value we can get a better performance. In other words this threshold governs the performance and we leave it to the wisdom of system administrator to fix an appropriate value for this threshold.

2) Real Application Dataset:

Similar to the last experiment; we tabulate the connection flows with standard deviation up to 10 in this case too. Table 5. shows periodic flows identified by the algorithm. First two entries in this case are two true positives corresponding to the sport web page and remaining are false positives. Although the number of false positives reduced compared to the previous case but this dataset represent less hours of activity in comparison to the previous one, thus reducing the traffic related to searches which were major contributors of false positives.

4 Conclusion

In this paper a data summarization problem is formulated for communication periodicity detection and an efficient implementation technique is described. Proposed method is a lossless summarization technique and is evauluated with traffic of simulated and real application having periodic communications.

References

1. `http://www.tcpdump.org`
2. Bartlett, G., Heidemann, J., Papadopoulos, C.: Using low-rate flow periodicities for anomaly detection: Extended. Technical report, University of Southern California (2009)
3. Chandola, V., Kumar, V.: Summarization- compressing data into an informative representation. Knowledge of Information Systems 12(3), 355–378 (2007)
4. Collins, M.: A Protocol Graph Based Anomaly Detection System. PhD thesis, School of Electrical and Computer Engineering, Carnegie Mellon University (2008)
5. Ertz, L., Eilertson, E., Lazarevic, A., Tan, P., Kumar, V., Srivastava, J.: The MINDS- Minnesota Intrusion Detection System, ch. 3 (2004)
6. Hubballi, N., Biswas, S., Nandi, S.: Towards reducing false alarms in network intrusion detection systems with data summarization technique. Security and Communication Networks 6(3) (2013)
7. Kim, M., Kang, H., Hong, S., Chung, S., Hong, J.: A flow-based method for abnormal network traffic detection. In: IEEE/IFIP Proceedings of the Network Operations and Management Symposium, pp. 217–228. Springer (2004)
8. Mahoney, M.V., Chan, P.K.: PHAD: Packet Header Anomaly Detection for identifying hostile network traffic. Technical report, Florida Institute of Technology (2001)
9. Mampaey, M., Vreeken, J.: Summarizing categorical data by clustering attributes. Data Mining and Knowledge Discovery 26(1), 130–173 (2013)
10. Mielikainen, T.: Summarization Techniques for Pattern Collections in Data Mining. PhD thesis, University of Helsinki (2005)
11. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: A new data clustering algorithm and its applications. Data Mining and Knowledge Discovery 1(2), 141–182 (1997)

Precedence Mining in Group Recommender Systems

Venkateswara Rao Kagita, Vineet Padmanabhan, and Arun K. Pujari

School of Computer and Information Sciences,
University of Hyderabad, India
venkateswar.rao.kagita@gmail.com, {vineet,akp}cs@uohyd.ernet.in

Abstract. We extend the `Precedence mining` model for *personal recommendation* as outlined in Parameswaran et.al., [6] in three different ways. Firstly, we show how precedence mining model can be used for recommending items of interest to a `group` of users (*group recommendation*) and compare and contrast our model with traditional group recommendation models like *collaborative* and *Hybrid*. Secondly, we extend the precedence mining model to incorporate `ratings` for items and experimental results show that the *goodness* of recommendation is improved. The third extension is related to the issue of *new items* being ignored which is a fundamental problem plaguing collaborative and precedence mining algorithms. When recommendations are based on other users interests (like in Collaborative recommender systems) the possibility of not recommending a new item which has not been consumed by many of the users is high though the new item may be of interest to the target user. We outline two models, `Vector precedence-mining` and `Hybrid precedence-mining` that addresses this issue.

1 Introduction

Collaborative filtering [2] is one of the commonly used methods for developing personal as well as group recommender systems [1,8,5,3,4]. Two main techniques used in collaborative filtering are called *Memory-based* and *Model-based*. Recently, Parameswaran et.al., [6] proposed a precedence mining approach for personal recommender system. The nicety and novelty of this approach is the use of pairwise precedence relations between items. Most of the earlier approaches of Collaborating Filtering do not take into consideration the relationship between objects. Though precedence mining, as demonstrated by Parameswaran et. al., [6], can be an important approach for recommender systems, current research is mostly limited to personal recommendation.

In this work we propose a group recommender system based on precedence mining model. Based on the precedence relations between pairs of items, a precedence probability of consuming an item by a user is calculated. The item with highest precedence probability for a user is recommended in a personal recommender system. We extend this concept of precedence probability to a group by introducing a virtual user. A virtual user, in principle, represents all users of the

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 701–707, 2013.

group. We derive the precedence probability of the virtual user of a group from precedence probabilities of individual members of the group. The group recommender system is extended further by incorporating users' ratings of items and by considering recommendation of unknown (new) item. The group recommender methods, proposed here, are experimented with real-life data on a movie database and is demonstrated that the user's satisfaction is substantially higher than earlier techniques.

2 Recommender Systems

The problem of recommender system is defined as follows. Let $O = \{o_1, o_2, \ldots, o_n\}$ be the set of items and $U = \{u_1, u_2, \ldots, u_m\}$ be the set of users. $profile(u_j)$ is the sequence of items consumed by user u_j. For given O, U, k and user $u \in U$, the Personal Recommender Problem (PRP) is to recommend k items to be consumed by user u. The recommended k items are presumed to be absent in $profile(u)$. The Group Recommender Problem (GRP) is to recommend k items to a group $G \subseteq U$ for given O, U and k. Any recommender system (personal or group) aims at selecting items for recommendation such that these items are expectedly preferred to other items by the user for whom it is recommended. Identifying most probable items from the set of profiles for a group of users (GRP) is harder than the same for individual users (PRP).

In general, collaborative filtering techniques find users having similar profiles as $profile(u)$ and then restricts its search to items consumed by this subset of users and not consumed by u. Thus certain patterns of consumption of items exhibited by the whole set of users, U, is not captured as the search is restricted. Precedence mining gets over this shortcomings and attempts to capture pairwise precedence relation occurring frequently among all users. Precedence mining approach as proposed by Parameswaran et.al. [6] is limited to recommending items for individual users. We, in this paper, propose a novel technique of use of precedence mining for group recommendation.

2.1 Precedence Probabilities

Given an item o_i, $support_i$ is the number of users who consumed item o_i. We define p_{ij} as the number of users having item o_i preceding o_j in their profiles. The precedence probability for items o_i and o_j, denoted as $PP(o_i|o_j)$ represents the probability of o_i preceding o_j. We define $PP(o_i|o_j) = \frac{p_{ij}}{support_i}$. There may arise cases where p_{ij} is zero and hence $PP(o_i|o_j)$ becomes zero. In order to avoid this situation we perturb PP by adding 1 to both numerator and denominator. Thus when p_{ij} is zero, we have $PP(o_i|o_j) = \frac{p_{ij}+1}{support_i+1}$. Let CI_j be the set of items consumed by user $u_j \in U$. Then we define

$$Score(o_i, u_j) = const \times \frac{support_i}{n} \times \prod_{o_l \in CI_j} PP(o_l|o_i) \tag{1}$$

For a group $G \subseteq U$, a trivial way of computing the score is to aggregate scores of individual members [9,7]. We propose a novel method of virtual user strategy. A virtual user $v(G)$ represents the whole group G. $profile(v(G))$ is generated from $profile(u_j), u_j \in G$. For group G, the set of consumed items CI_G is $\underset{u_j \in G}{\cup} CI_j$. We define weight for an item consumed by an individual user as

$$weight(o_i, u_j) = \begin{cases} 1 & \text{if } o_i \in CI_j \\ score(o_i, u_j) & otherwise \end{cases}$$

We define weight of an item for group G as

$$weight(o_i, G) = \sum_{u_j \in G} weight(o_i, u_j)/|G|$$

An item $o_i \in profile(v(G))$ if $weight(o_i, G) \geq \tau$ where τ is a predefined threshold. Algorithm 1 gives pseudo-code of constructing $profile(v(G))$.

Algorithm 1. Virtual User Strategy in Probabilistic Model

Input: O, U, $profile(u, \forall u \in U)$, G, τ
Output: Recommended Item(s)
$profile(v(G)) = \emptyset$;
for *each item* $o_i \in CI_G$ **do**
  Calculate $weight(o_i, G)$;
  if $weight(o_i, G) > \tau$ **then** $profile(v(G)) = profile(v(G)) \cup o_i$;
  delete o_i from CI_G
for *each item* $o_i \in O \setminus CI_G$ **do** Calculate $score(o_i, v(G))$;
Recommend top k items that maximizes $score(o_i, v(G))$.

2.2 Precedence Mining Model with Ratings

In many real life situations a user not only consumes an item, but also gives a satisfaction rating for the item. A group recommendation system becomes more meaningful if such ratings are also incorporated in the user profile. We assume here that $profile(u_j)$ is a sequence of ordered pairs (o_i, r_i) indicating the item o_i consumed by user with a rating of r_i. In this situation we propose a new method of computing the precedence probability for a pair of items. There is a lower bound t such that cases with $r_i \geq t$ are only considered. We define $p_{ij}(r, t)$ as number of user profiles in which item o_i precedes item o_j with $r_i = r$ and $r_j \geq t$. Similarly $support_i$ is number of user profiles that contain item o_i with rating $r_i \geq t$. $PP(o_i : r|o_j) = \frac{p_{ij}(r,t)}{support_i}$ and

$$Score(o_i : r, u_j) = const \times \frac{support_i}{n} \times \prod_{o_l \in CI_j} PP(o_l : r|o_i) \tag{2}$$

In Parameswaran et.al., [6] incorporating ratings into the precedence mining model was left out as part of future work.

3 Vector Precedence Mining

Majority of the group recommender systems try to recommend an item o_i belonging to the set of known items O. There are situations when a new item is introduced in the market and it becomes necessary to consider recommending this new item. One of the drawbacks of the existing precedence mining models is their inability to incorporate new items. To handle such situations, $profile(u_j)$ is extended as a sequence of vectors representing multiple features of items. Thus instead of counting the number of occurrences of an item, the number of occurrences of feature-set is counted. Therefore an unknown item which is similar in features to any of the known items could be treated as an identical item. For the case of movies, each movie is represented as a set of genres like *Action, Comedy* etc.[1] $profile(u_j)$ is a binary vector with 1 if a genre is present in the movie, otherwise zero. We take the decimal equivalent of that binary vector and represent the movie with that value.

3.1 Social Influence

Many persons, often, like an item because it is associated with persons they like. In the case of a movie recommendation scenario, people watch movies of their favorite star (hero, heroine, director, etc.,) irrespective of the content of the movie and they also watch movies of the star who is having similar characteristics or relevance to their favorite star. It is important to consider such *social influence factors* in a group recommender system. Our experimental results show that the *goodness* of recommendation is improved while taking into consideration such social factors with respect to a movie dataset. In our experiments, we augment $profile(u)$ by taking into consideration factors like actor-id, director-id etc.

3.2 A Hybrid Technique Based on Precedence Mining

In section3 we outlined a vector representation model to overcome the issue of new items not being recommended to the target user/s. We propose a hybrid technique that combines both precedence mining and content based recommendations. Our idea is as follows. First apply precedence mining technique and get top $2k$ (where k is the number of items to be recommended) items. Having done that get score for new items. Then apply content based technique and get score for these top $2k$ and new items. Add these two scores and re-rank according to the new score.

4 Experiment Results

The concepts introduced in the above sections are validated in a movie recommendation scenario. Our dataset consists of 100 users and 453 movies. User profiles with ratings were created by consulting these individuals. Our recommender

[1] In the data-set used for our experiments each movie is represented with 19 genres.

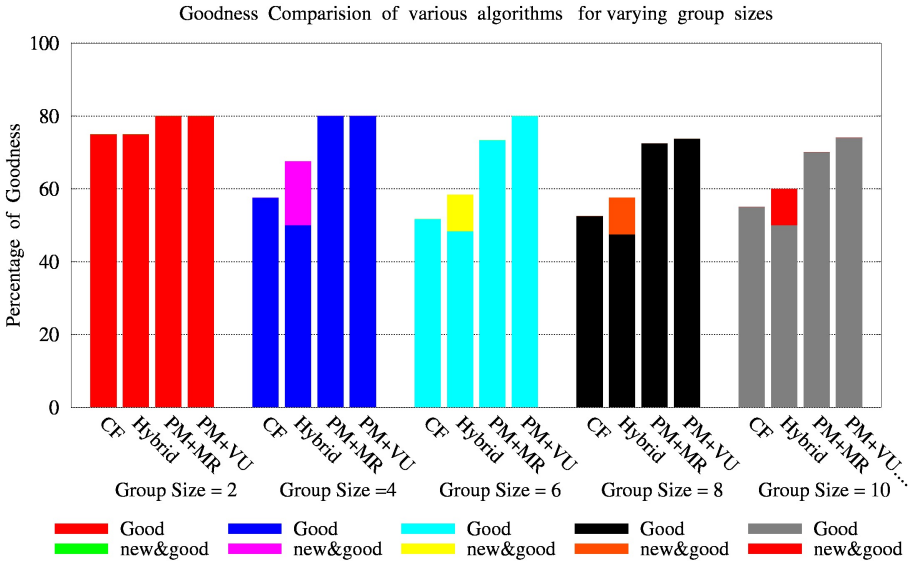


Fig. 1. Goodness comparison Collaborative and Hybrid Vs. Precedence Mining

algorithms are validated by collecting users' satisfaction on the outcomes of our algorithms as well as on the outcomes of collaborative and Hybrid algorithms. We presented the recommendations of all the algorithms (each movie only once) and asked them to give their satisfaction on that recommendation in a 1-5 scale. (Note here that the satisfaction may depend on many parameters such as theme of the movie and surprise about recommendation and many more). User/s satisfaction rating exceeding 4 are only considered to calculate the percentage of goodness as shown in equation 3 wherein $response_{ij}$ is response given by user $u_i \in G$ for movie m_j

$$percentage\ of\ goodness = \frac{\sum_{i=1}^{|G|} \sum_{j=1}^{k} response_{ij} \geq 4?1:0}{k \times |G|} \times 100 \qquad (3)$$

First set of experiments are to compare the base models of precedence mining with the traditional collaborative and hybrid techniques for different group sizes. Figure 1. shows the results wherein CF represents collaborative filtering, PM+MR represents precedence mining with merging results strategy and PM+VU represents precedence mining with virtual user strategy. We observe that in Figure 1. newness is completely missing in collaborative and precedence mining models. There is an overall improvement of goodness in precedence mining models compared to collaborative and hybrid models. PM+VU has consistently better performance than PM+MR. In order to calculate *new&good* movies we consider only recently released movies from the recommendation set whereas

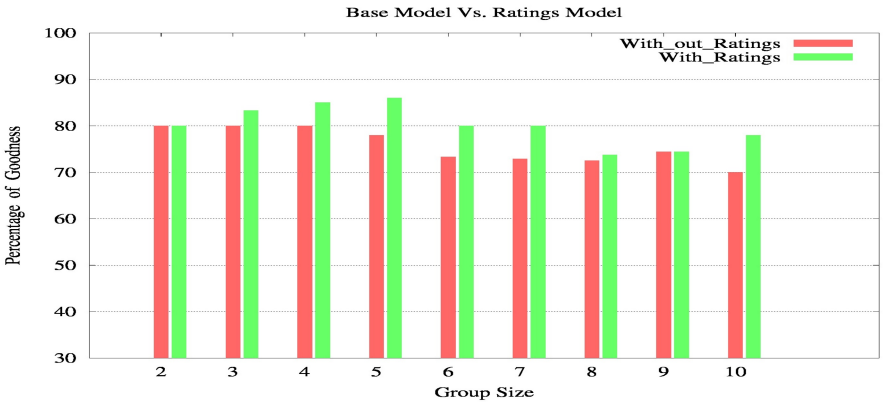


Fig. 2. Performance difference with and without incorporating ratings

to calculate *Good* we use the remaining movies in the recommendation set. The second set of experiments are related to the improvement of goodness by incorporating ratings into the precedence mining model (as described in section 2.2). This set of experiments are carried out by averaging the individual scores and not by virtual user strategy. Figure 2. shows these results comparison. If we observe

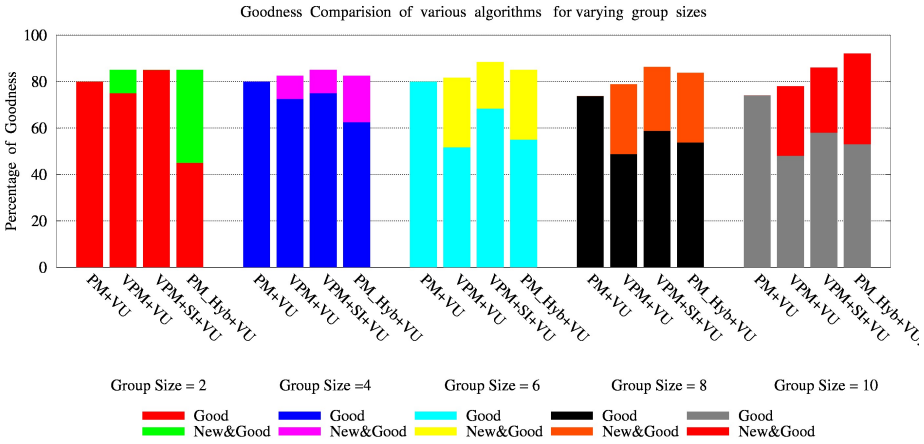


Fig. 3. Performance Improvement with vector precedence mining and social influence

the graph in Figure 2. incorporating ratings into the precedence mining model is performing better than without ratings most of the time or it is guaranteed that incorporating ratings is at-least as good as without incorporating ratings. If we observe Figure 1., Virtual User strategy is performing better than Merging Results strategy in most of the cases. So we have used virtual user strategy in vector-precedence mining model and social influence model as outlined in section 3. Results related to `newness` improvement with vector-precedence mining

and `goodness` improvement by considering social influence factors such as *actor* in movie domain (described in section 3.1) is shown in Figure 3. VPM+VU in this figure represents vector precedence mining with virtual user strategy, VPM+SI+VU represents vector precedence mining and social influence with virtual user strategy and PM_Hyb+VU represents precedence mining hybrid model with virtual user strategy. The dataset related to our experimental results can be found at `http://dcis.uohyd.ernet.in/~vineetcs`.

5 Conclusions and Future Work

In this work we have developed a framework for group recommender systems (GRS) based on precedence mining. We outline two models which has the advantage of extracting more frequent patterns within the whole set of user profiles as compared to other GRS based on collaborative filtering. Our experimental results show that both *goodness* as well as *newness* of recommendation is improved using our group recommender algorithms. In the future, we would like to extend our framework to include rating prediction for virtual user items as well as to provide justification for group recommendations.

References

1. Amer-Yahia, S., Roy, S.B., Chawla, A., Das, G., Yu, C.: Group recommendation: Semantics and efficiency. PVLDB 2(1), 754–765 (2009)
2. Breese, J., Heckermanx, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: UAI (1998)
3. Chen, Y.-L., Cheng, L.-C., Chuang, C.-N.: A group recommendation system with consideration of interactions among group members. Expert Syst. Appl. 34(3), 2082–2090 (2008)
4. Padmanabhan, V., Kiran, P., Sattar, A.: Group recommender systems: Some experimental results. In: 5th International Conference on Agents and Artificial Intelligence (ICAART 2013), Barcelona, Spain, pp. 370–376 (2013)
5. Padmanabhan, V., Seemala, S.K., Bhukya, W.N.: A rule based approach to group recommender systems. In: Sombattheera, C., Agarwal, A., Udgata, S.K., Lavangnananda, K. (eds.) MIWAI 2011. LNCS, vol. 7080, pp. 26–37. Springer, Heidelberg (2011)
6. Parameswaran, A.G., Koutrika, G., Bercovitz, B., Garcia-Molina, H.: Recsplorer: recommendation algorithms based on precedence mining. In: SIGMOD Conference, pp. 87–98 (2010)
7. Seko, S., Yagi, T., Motegi, M., Muto, S.Y.: Group recommendation using feature space representing behavioral tendency and power balance among members. In: RecSys, pp. 101–108 (2011)
8. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Providing justifications in recommender systems. IEEE Transactions on Systems, Man, and Cybernetics, Part A 38(6), 1–1272 (2008)
9. Yu, Z., Zhou, X., Hao, Y., Gu, J.: Tv program recommendation for multiple viewers based on user profile merging. User Modeling and User-Adapted Interaction 16(1), 63–82 (2006)

Optimizing Research Progress Trajectories with Semantic Power Graphs

G.S. Mahalakshmi[1] and S. Sendhilkumar[2]

[1] Department of Computer Science and Engineering, Anna University
[2] Department of Information Science and Technology, Anna University
{gsmaha,thamaraikumar}@annauniv.edu

Abstract. Any researcher who is taking up a new research work must explore the works done in the past. For this we propose an idea to track the possible work progresses of a particular research article through semantic based approaches. In addition we analyze the co-citations and cross-citations among research works to avoid leaving out any significant works of the past. Finally we attempt to represent the citation networks and related meta-info using power graphs. This technique reduces the overhead of huge dimensions of citation networks thereby providing optimized trajectory representations which leads to finding significant research progress trajectories.

Keywords: citation, co-citation, trajectory, semantics, H-index, power graph, main path, key-route path, backward ideal path.

1 Introduction

An Integrated Approach for Tracking Work Progress is a bibliometric method capable of tracking the most significant paths of work progress and is also used to trace the developing trajectory of a research paper. The study provides valuable information for academic researchers and scholars to gain insights into development trajectory of the given research paper. It incorporates the integrated main path analysis approach, with which it helps to view the true path in which the research work has progressed. The outcome of the study shows the most significant work progress graph of the research paper. It also captures the works that are carried out in parallel and country contribution.

Many researchers have applied the main path analysis method since then to investigate the developmental trajectories of various science and technology domains using bibliographical citation data, patent citation data, or both [2, 4-8, 11]. Hummon and Doreian [9] addressed the same issue by introducing main path analysis. The method helps to escape from the maze by offering the most significant trajectories the main paths of a citation network. Methodologically, Batagelj [1] proposed a major advancement of the method, enhancing main path analysis by offering efficient algorithms for determining various versions of the significance index.

An earlier work on patent data [4] integratse social network analysis and main path analysis to explore the technological development trajectories of automatic

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 708–713, 2013.

identification techniques. The method, however, has some limitations, as it either offers a complex "network of main paths," as suggested in Hummon et al. [10], or one and only the most significant development path. For a large citation network, the "network of main paths" does not achieve the goal of simplifying the citation network. On the other hand, the one and only path, which limits our view to the target science or technology domain, is not satisfactory for explorers who are looking for more than the most significant development path. Further, the "priority first search" algorithm traces the most significant route at each juncture when several new ideas are competing. The result obtained may not be the path with the largest overall impact. Indeed, many close-second significant routes also can be neglected in the process.

In this paper, we attempt to apply power graphs over the semantically built citation network through which the 'bigness' of citation networks is reduced to a greater extent. We analyse the citation networks to evolve technological work progress which are semantically appealing unlike the earlier methods. We have suggested three different optimization approaches in finding main path analysis over semantic citation networks.

2 Forming Work Progress Trajectories

The citations of the seed paper are denoted as N, at the first generation. All the immediate first level citation papers are retrieved for the seed paper. The procedure is iterated to retrieve all the citation papers at all generations till 2012. Next to this, the corpus is subjected to preprocessing.. After these steps similarity between the text document are found using cosine similarity and topics are found using LDA algorithm. The cosine similarity value is determined for comparing seed paper and the citation papers. The average value of every level is also determined. Papers with similarity score greater than the average are only retained and the rest are eliminated. This dynamic threshold value is used to eliminate citation papers that do not semantically match with the seed paper. Citation papers that are semantically related to the seed paper only survive after this phase. Rest of the papers is filtered. The set of citation papers that are preserved after elimination is notated by N_p, where $N_p \leq N$. This is done for all levels.

Next we proceed to find out the co-citations of papers and trace all possible co-citations. To find the co-citations we used Cite Seer. This enables us to find the co-citations and compute the bibliographic coupling strength. For all the citation papers we determine the co-citations which is notated as M_{ij}, where i is a co-citation to j, i= 1 to n, j=1 to k where k is the no of co-citations in N_i. For all the co-citation papers, we perform the semantic analysis, get their concept maps and compare with the seed paper to generate a semantic index score and use a dynamic threshold value to determine the semantically matched citation papers. The set of co-citation papers that are preserved after elimination is notated by $N_{co\text{-}p}$.

Graph Node blending and Feature preservation are done in redundancy elimination phase. Level 2 Graph (G2) is an input to the redundancy elimination phase. The graph formed in level 2 will have redundancies. The graph may contain duplicate nodes which have to be removed. The graph may or may not contain duplicates. If duplicates are present Graph Node Blending has to be done. The papers which are

present in N and not present in N_p and has again appeared in $N_{co\text{-}p}$ whose features has to be preserved. Has a Output of this module we get Reduced (N_p,$N_{co\text{-}p}$). After redundancy elimination we get Reduced (N_p,$N_{co\text{-}p}$) set of papers for which cross citation analyses have to be performed (refer section: Results). For instance, if a paper occurs in citation levels,L2,L4,L7 in graph(G1) and the same paper occurs in co-citation levels CO1 and CO3 in the graph(G2),then the node is blended with all these information as L2,L4,L7,CO1,CO3 in G2.

We further proceed to include cross citations between papers in our corpus across the entire timeline. Cross citation can occur across levels as well. We use k-means clustering and link analysis for finding the cross citations. Clustering is done to group the papers according to the topics dealt in the papers. In each cluster the new citation links between the papers are found and results in increase of edges in the graph (G2), which leads to the graph (G3).

We construct a graph by considering the seed paper as the root node with the citation papers N_p as the child nodes with directed edges indicating the citation relationship with the seed paper. This is done at all generations. We form an acyclic graph that has directed edges and is devoid of cycles, initially to start the graph formation process. Nodes represent the papers and the edges represent the citation relationship between them. In G1, the nodes purely represent the citation papers whereas in G2, the nodes include both the citation and co-citation papers. In G2,there is an increase of nodes and edges present whereas in G3, no new nodes are added, but results in increase of edges.

1. Directed Acyclic Graphs (DAG's): We obtain the first graph G1 after semantic analysis with filtration using DAG. The citations of the seed paper at all generations are represented using directed acyclic graphs. Constructing DAGs ensures there are no cycles within the graphs. Consider C1, C2, C3 to be the citation papers of the seed paper, S. Let C1 be present at levels L1,L3 and C2 to be present in levels L4 and L5.Let C3 belong to level L2.These are represented in DAG using 5 citation nodes and a seed paper node. We call this graph, say G1.

2. Power Graphs: Power Graphs are used to reduce the dimensionality in graphs. C1 node can be blended with the information that it occurs in levels L1, L3 in a single powered node. Similarly the other citation papers C2 and C3 are represented using power nodes. Now the graph G1 can be represented using 3 citation nodes and a seed paper node, thus reducing dimensionality. The two conditions to be satisfied for a power graph are:
    - Power node hierarchy condition: Any two power nodes are either disjointed or one is included in the other.
    - Power edge disjointness condition: There is an onto mapping from edges of the original graph to power edges

Since we are representing the nodes as power nodes, they have information about which citation and co-citation levels it occurs. So each node is represented using multi colors, which denote the levels it occurs. The co-citations for the graph G1 is also brought into the graph representation. Since this involves increase in number of

nodes, the co-citation information is also blended within the power nodes. If M1 and M2 are co-citations of C1, 2 co-citation nodes are to be added to G1. But this co-citation information is blended in C1 power node thus bringing in co-citation information without adding further nodes. We call this graph as G2. The citations that are present between various clusters are analyzed in cross citation analysis. Suppose if C1 belongs to network cluster and C2 in software engineering cluster. The citations between them are analyzed after clustering the citation papers. After bringing in cross citation information in G2, we call the final graph as G3.

We use four approaches to track the work progress of a paper in the integrated approach. They are namely: Global Main Path, Backward Local Main Path, Multiple Main Path, and Key Route Main Path. The global main path searches forward from the source to the sinks and is traced using Priority First Search. This uses the Dijkstra's Algorithm, which finds the path with the lowest cost. The Backward Local Main Path is found using Reverse Priority First Search Algorithm. We trace the path backwards according to semantic scores from a recent paper until we reach the seed. For the nodes that are retained through optimization approach 1, we find the paths from leaf node towards seed node. Multiple Main Path uses Semantic Index Relaxation which basically relaxes the search constraint. The constraints are relaxed by bringing the next longest path also. Using the Key route Algorithm we take the most significant link and begin a search from the key rather than from the source or sink. This key is the research paper with a very close semantic score compared to the seed. It guarantees that this key route is included in the main path.

3 Optimising Work Progress Trajectories

The Paths obtained are ranked based on path length, node popularity, and relevancy with the seed paper. More Length indicates that more frequently the continuing research works have been carried out. Generally any significant research idea attracts more citations, hence is an indicative of good work progress (fetched through high citation count). The citation counts for all the nodes in a path are added and the average is taken as popularity score. Average of similarity scores of nodes in a path with respect to the seed paper, gives relevancy score which indicates more relevant work progress. The Paths found in the graph (G3) is quite large in number. Hence in order to improve the quality of the paths found we propose three optimization approaches and the results are discussed.

Approach 1: Matching Paths by Semantic score Average Filtration Level wise
In this approach, the semantic score for each node in the graph is determined from the cosine similarity algorithm in comparison with the seed paper. The nodes in each level are considered separately and their semantic score average is calculated for each level and the nodes whose semantic score are above each level's average are only considered and retained. Later paths are determined for the filtered nodes that offers better semantic match between the papers in the path and reduction in the number of paths between any nodes.

Approach 2: Using Semantic score Average Filtration Level wise
In this approach semantic score is considered for tracing the path from the seed till the leaf. It is a very narrow approach in which we consider only the nodes that survive the threshold based semantic filtration of first level. Next level nodes are considered only for the nodes that have survived the previous level filtration. Since the corresponding nodes of prior filtered nodes are alone considered, the number of nodes under optimization is reduced largely in number. It proceeds till the latest level generation, which is fourth level. The difference between approach 1 and approach 2 is that, in the first approach the filtration was done level wise before and then the link was found whereas in the second approach only the nodes which have survived the first level filtration are considered for the second level and hence forth. The path obtained thus obtained is narrow.

Approach 3: Using Semantic score and Popularity Average Filtration Level wise
In this approach we use two criteria's namely semantic score and popularity of the paper for narrowing the graph. The similarity between all papers of the first level is determined using cosine similarity algorithm. The average semantic score for the first level alone is taken into account and the papers that are above the average are retained. For the remaining levels we use the popularity score as criteria for reducing the paths. We take the average of the popularity for the second level and for all the remaining levels. The papers above average are retained and thus results in a fewer number of paths than before optimization.

Approach 4: Best Semantic Match Path
In order to find the best semantic match path, for all the nodes in the graph we find the cosine similarity score with the seed paper. Then we find the nodes with maximum semantic score at each levels. The best semantic path is the path which contains these maximum semantic score nodes.

4 Results and Discussions

The papers across seven generations are taken and a link analysis is made across papers. The seed paper [3] has totally 2706 and 2038 citations at the first and second generations. After the semantic scores are computed a dynamic threshold based filtration is performed. The optimized research progress paths are listed in Table 1.

Table 1. Optimization Approaches

Approach	Paths	From Node – To Node
Optimization I	26 Paths	000 - 56
Optimization II	000-174-167	000 - 180
Optimization III	000-174-167	000 - 167
Best	000-209-102-141-168	000 - 168

5 Conclusion

We have proposed semantics based approach to track the work progress of a research paper across timeline. We include co-citations and cross-citations to improvise the graph at every stage to obtain an efficient graph. We have also proposed different optimization methods to get better results. In the future, the download can be automated.

References

1. Batagelj, V.: Efficient algorithms for citation network analysis. Prep. Ser. Univ. Ljubljana, Inst. Math. 41(897), 1–29 (2003)
2. Carlero-Medina, C., Noyons, E.C.M.: Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. Journal of Informetrics 2(4), 272–279 (2008)
3. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences of the United States of America 102(46), 16569–16572 (2005)
4. Lu, L.Y.Y., Lan, Y.L., Liu, J.S.: A novel approach for exploring technological development trajectories. In: 2012 IEEE International Conference on Management of Innovation and Technology (ICMIT), pp. 504–509 (2012)
5. Lu, L.Y.Y., Lin, B.J.Y., Liu, J.S., Yu, C.Y.: Ethics in nanotechnology: What's being done? What's missing? Journal of Business Ethics 109(4), 583–598 (2011)
6. Lucio-Arias, D., Leydesdorff, L.: Knowledge Emergence in Scientific Communication: From "Fullerenes" to "Nanotubes". Scientometrics 70(3), 603–632 (2007)
7. Mina, A.R., Ramlogan, G., Tampubolon, J.S.: Metcalfe,'Mapping Evolutionary Trajectories: Applications to the Growth and Transformation of Medical Knowledge'. Research Policy 36(5), 789–806 (2007)
8. Moore, S., Haines, V., Hawe, P., Shiell, A.: Lost in translation: A genealogy of the "social capital" concept in public health. Journal of Epidemiology and Community Health 60, 729–734 (2006)
9. Hummon, N., Doreian, P.: Computational methods for social network analysis. Social Networks 12, 273–288 (1990)
10. Hummon, N.P., Doreian, P., Freeman, L.C.: Analyzing the Structure of the Centrality-Productivity Literature Created Between 1948 and 1979. Science Communication 11(4), 459–480 (1990)
11. Verspagen, B.: Mapping Technological Trajectories as Patent Citation Networks: A Study on the History of Fuel Cell Research. Advances in Complex Systems 10, 93–115 (2007)

An Induced Fuzzy Rasch-Vikor Model for Warehouse Location Evaluation under Risky Supply Chain

Kajal Chatterjee* and Samarjit Kar

Department of Mathematics, National Institute of Technology,
Durgapur-713209, West Bengal, India
{chatterjeekajal7,dr.samarjitkar}@gmail.com

Abstract. This paper addresses dynamic risky nature of supply chain in warehouse site evaluation where the target is to design network flow of products so that all customers demands are satisfied at minimum risk in distribution network. Main theme of the paper is to choose the most optimal and low risk warehouse spot from a number of potential alternatives locations. A new Fuzzy- Rasch-VIKOR decision model is provided where uncertain information are exploited getting group of decision makers as multiple experts providing decision through triangular fuzzy numbers. A two-phase algorithmic approach is proposed to deal with the problem. First phase involves identification of warehouse location, evaluation of risk criteria using fuzzy Rasch model to quantify criteria weights under uncertainty and second phase ranks the location alternative by fuzzy- VIKOR method for selecting the optimal low risk site. Finally we demonstrate our decision model with a case study illustrating the application in risky supply-chain.

Keywords: Fuzzy Rasch Model, VIKOR, Multiple criteria decision making (MCDM), Supply chain risk management (SCRM).

1 Introduction

Evaluation of warehouse location in a complex supply chain is a typical multiattribute based decision making problem having various risk criterias.A poor choice of location may result in lower logistic performance, minimal usuage of resources and loss of competitive advantage in market. In supply chain management, warehouse function as consolidation points where products shipments from production sites are assembled or sorted. In supply chain warehousing selection, many methods were proposed by researchers in past years. Chen [1] presented a fuzzy MCDM problem for location of distribution center. Hidaka and Okano [2] determined a simulation method for selection of hefty warehouse location. Michel and Hentenryck [3] put forward a tabulated search algorithm using linear neighborhood on incapacitated warehouse location problem. Sharma and Berry [4]

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 714–719, 2013.

gave an idea of locating single-stage computed warehouse. Opricovic [5] developed complex multi-criteria optimization using Ideal point VIKOR. Andrich [6] developed Rating Scale Model and adopted five-point Likert scale in parameter evaluation and generating fuzzy numbers. Yu and Wu [7] utilized the Rasch Item Response Theory to evaluate fuzzy elements. Huang and Peng [8] combined Rasch model, fuzzy logic and TOPSIS method to access competition in Asian region tourism sector. This study proposes a new hybrid model, the fuzzy Rasch model in VIKOR combining Rasch model, fuzzy theory and VIKOR, which is discussed below. The above model is applied in warehouse selection for a global electrical manufacturer company, from among eight location alternatives in India taking into account various risks in supply chain.

2 Basic Concepts

2.1 Rasch Model

Rasch Models, also known as One-Parameter Logistic and static model under Item response theory(IRT) models in field of psychometry quantifying human behaviour. Rasch Model was developed for two choice items. It models the relationship between person's ability and item difficulty as probabilistic function. Correct data are obtained based on degree of latent proficiency function. Measuring in logits(log odd units) the model converts initial raw data from rating scale to equal interval scale showing both dificulty of item and individual ability. Rasch model provides the fuzzy perception logic model with precise properties far comprehensive to fuzzy logic.

2.2 Rating Scale Model

Andrich[6] who developed Rating scale model(RSM), is a additive linear model describing probability of response of specific person(n) to specific Likart item(i)with specific rating scale step(x). RSM applies Rasch model to polytomous rating scale instruments,including five-point Likart scale. RSM is applied to assess the values of parameters, shown as: $log\{P_{nij}/P_{ni(j-1)}\} = \theta_i - (\delta_i + \tau_j)$ where P_{nij} = probability the item n obtain j scores from expert i; $P_{ni(j-1)}$ = probability the item n obtain j-1 scores from expert i;θ_i = item difficulty of item n; δ_i = individual ability of expert i; τ_j = step difficulty of category j.

2.3 Vikor Model

The VIKOR[5] (VIekriterijumsko KOmpromisno Rangi ranje) method was proposed by Opricovic for multi-criteria based methodology in complex Network process based on ideal point method. It evaluates the ranking list, comprehenive solution and weight stability intervals for preference alternatives in respect to initial (given) weights. Based on ranking index it determines closeness of given data to ideal solution. By introducing multi-criteria index based

on measure of closeness to ideal solution compromising ranking list and solution are evaluated. VIKOR method developed in form of L_p metric. $L_{p,i} = \{\sum_{j=1}^{n}\{w_j(f_j^* - f_{ij})/(f_j^* - f_j^-)\}^p\}^{1/p}, 1 \preceq p \preceq \infty, i = 1, 2, ...m$. In VIKOR model, $L_{1,j}$(asS_i)and $L_{\infty,j}$(as R_i) are applied in ranking measurement.The solution gained by min S_i and by min R_i is with a maximum group utility of the majority, and minimum individual regret of opponent respectively.

3 Algorithmic Approach to Fuzzy-Rasch Based VIKOR Model

For the function of risk criteria in selection of warehouse location, our proposed model uses Fuzzy Rasch method to propagate the triangular fuzzy weights of the risky criteria. Then defuzzified fuzzy weights are then applied in Fuzzy VIKOR for ranking the optimal warehouse location zones. The algorithmic steps are as follows:

Step 1: Risk criteria and the corresponding alternatives are evaluated based on decision makers judgment.

Step 2:Based on Fuzzy Rasch Model the weights of the risk criteria are generated.

2.1 Step parameters (δ_{ij}) for determining fuzzy weight are calculated based on decision-makers importance for each criterion. Using triangular fuzzy weight and arithmetic average to evaluate the fuzzy weight of each expert,$\tilde{W}_C = (\delta_c^L, \delta_c^M, \delta_c^U)$.

2.2 The fuzzy criteria weights are transform into crisp weights through defuzzification, done by best on Non-fuzzy performance value, BNP_c= $(\delta_c^L + \delta_c^M + \delta_c^U)/3$, which are used as criteria weights in VIKOR. Based on above, this study standardize the weights W_c of criterion, $SBNP_c$; $W_c = SBNP_c = BNP_c / \sum_{c=1}^{c} BNP_c$ where c represent the number of criteria.

Step 3: A decision matrix is formed based on calculated values of alternatives with respect to each criteria. The defuzzified values of criteria to the alternatives are calculated.

Step 4: Determine the best f_i^* and worst f_i^- values of all criteria functions, for i^{th} risk criteria function, $f_i^* = min_j f_{ij}$ and $f_i^- = max_j f_{ij}$.

Step 5:Compute the values of S_j, R_j and Q_j, $j = 1, 2, ...J$ by the relations $S_j = \sum_{i=1}^{n} w_i(f_i^* - f_{ij})/(f_i^* - f_i^-)$, $R_j = max_j w_i(f_i^* - f_{ij})/(f_i^* - f_i^-)$, $Q_j = \frac{v(S_j - S^*)}{(S^- - S^*)} + \frac{(1-v)(R_j - R^*)}{(R^- - R^*)}$, $S^* = min_j S_j$, $S^- = max_j S_j$, $R^* = min_j R_j$, $R^- = max_j R_j$ where w_i = weights of criteria and v = weight of maximum group utility.

Step 6: Based on values of S_j, R_j and Q_j, $(j = 1, 2, ...J)$; final ranking of alternatives is proposed.

Step 7:Alternative $(a^{'})$ is ranked the best by measure Q (minimum): $Q(a^{"}) - Q(a^{'}) \succeq DQ$ where $(a^{"})$ is second best alternative by Q: $DQ = 1/(J-1)$ where J is the number of alternative and alternative $(a^{"})$, must also be best ranked by S or/and R. The best alternative, ranked by Q, is the one with minimum value of Q. The main ranking result is the compromise ranking list of alternatives.

4 Case Study

Havell's India ltd, a billion-dollar-plus power distribution equipment manufacturer organization has 94 branches and over 8000 professionals in over 50 countries across the globe. It plans to select new warehouse in India to improve its work productivity following ten risk Criteria are under consideration: Political Instability risk (C_1); Environmental risk (C_2); Market demand and sale risk (C_3); Logistics and cargo damage risks (C_4); Natural calamities risk (C_5); Inflation rate risk (C_6); Custom delays and damage risk (C_7); Information flow and affordability of service risk (C_8; Agreement terms and type risk (C_9); Change in product design and operation quality risk (C_{10}). Five expert decision-makers (DMs),D_1 (Electrical Engineer),D_2 (Marketing Manager),D_3 (Economics analyst),D_4 (Statisticians) and D_5 (Production Manager), are engaged for selecting and ranking low risk locations among eight alternatives, namely: Surat (A_1), Gurguon (A_2), Bangalore (A_3), Ahmadabad (A_4), Aurangabad (A_5), Nasik (A_6), Kanpur (A_7) and Hyderabad (A_8),with respect to these ten risk attributes. Following steps of the above algorithm are utilized and the results are shown in the Tables below:

Step 1: $X = (A_i; i = 1, 2, ..., 8)$ be a set of 8 alternative location for warehouse and $F = (C_j; j = 1, 2, 3, ...10)$ be a set of risk criteria's.
Step 2: Using Sub-steps 2.1-2.2, calculate the step parameters to generate the fuzzy weight and degree of importance of criteria accoring to experts, shown in Table1-2 to generate the weights of evaluation criteria, shown in Table 2.

Table 1. Expert DMs five point Likert scale based on Rasch Model

Expert	Very Unimp-	Unimportant	General	Important	Very Impor-
DM_j	portant (1)	(2)	(3)	(4)	-tant(5)
DM_1	-	(2,2,8)	(2,7,12)	(5,11,17)	(11,17,17)
DM_2	(0,0,6)	(0,3.5,7)	(3.5,8,12.5)	(8.5,13,17.5)	(13.5,18,18)
DM_3	-	(1.5,4,6.5)	(4.5,8,11.5)	(8,12.5,17.5)	(13,18,18)
DM_4	(1,1,5)	(2,4,6)	(3.5,7.5,11.5)	(7,12,17)	(12,17,17)
DM_5	-	(1,1,5)	(3,7,11)	(6.5,11.5,16.5)	(11,16.5,16.5)

Table 2. Importance of criteria according to five decision makers (DMj) and weight of criteria

Criteria	DM_1	DM_2	DM_3	DM_4	DM_5	Average Wt	Defuzzified	Standardise
						W_C	W_C= BNP_C	W_C= $SBNP_C$
C_1	4	4	5	4	5	(8.9,14.1,17.2)	13.40	0.13180
C_2	4	5	5	5	5	(10.9,16.1,17.3)	14.76	0.14524
C_3	4	4	3	4	4	(6.3,11.1,15.9)	11.10	0.10918
C_4	5	4	3	4	3	(6.8,11.4,14.8)	11.00	0.10819
C_5	3	2	4	5	4	(5.7,10.3,14)	10.00	0.09836
C_6	3	2	3	2	3	(2.3,5.9,9.5)	5.90	0.05803
C_7	2	3	2	1	3	(2.2,4.4,8.6)	5.06	0.04983
C_8	4	5	4	5	5	(9.9,15,17.2)	14.03	0.13803
C_9	2	4	3	2	4	(4.3,6.9,10.7)	8.10	0.07967
C_{10}	4	2	3	4	3	(3.9,8.3,12.7)	8.30	0.08163

Table 3. Triangular Fuzzy Scale of the Linguistic Variables

Linguistic Variables	Triangular Fuzzy values	Defuzzified Value
Poor(P)	(0,0,3)	1
Fair(F)	(2,5,8)	5
Good(G)	(6,10,14)	10
Very Good(G)	(11,14,17)	14
Excellent(E)	(15,18,18)	17

Step 3: Eight alternative are evaluated with respect to ten risk criteria and all values assigned to alternative in respect to each criteria form a decision matrix by help of Table 4. Also positve and negative ideal solution values of all criteria are shown in Table 5.

Table 4. Ratings of criteria to alternatives with best f_i^* and worst f_i^- value

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
A_1	G	E	F	P	F	E	F	P	E	VG
A_2	VG	F	VG	F	G	E	F	VG	E	P
A_3	F	VG	E	F	G	F	G	P	VG	F
A_4	E	G	F	E	G	G	VG	E	P	VG
A_5	F	E	P	G	VG	F	G	P	VG	P
A_6	P	E	VG	G	E	P	VG	F	F	G
A_7	G	VG	G	F	P	E	P	G	F	E
A_8	F	F	E	E	F	F	VG	P	F	VG
f_i^*	1	5	1	1	1	1	1	1	1	1
f_i^-	17	17	17	17	17	17	14	17	17	17

Step 4: Compute the values of S_i ,R_i and Q_i and rank the respective alternatives using fuzzy VIKOR model, shown in Table 5.

Table 5. Calculating the values of S_i ,R_i and Q_i and ranking of alternatives

Warehouse location (A_j	Group) Utility S_j	Individual Regret R_j	Aggregating Index Q_j	Rank of alt--ernatives (v=0.5)
1	0.490642	0.145246	0.645270	3
2	0.543367	0.112152	0.307156	5
3	0.510629	0.109180	0.193227	6
4	0.669986	0.138033	0.945935	1
5	0.467227	0.145246	0.593049	4
6	0.543361	0.145246	0.762848	2
7	0.508770	0.108934	0.185696	7
8	0.425512	0.109180	0.003393	8

5 Result Discussion

Using Step 7 of algorithm, $Q(A_4) - Q(A_6) = 0.945935 - 0.762848 = 0.183087 \succ 1/7$, so the first condition is satisfied. Also, alternative (A_4) also ranked best in terms of risks by corresponding values of S_j or Q_j considering $v = 0.5$ (for unbiased judgment of decision makers).

6 Conclusion and Future Areas of Research

The present study demonstrates the feasibility of applying Fuzzy Rasch method in VIKOR to analyze optimal low risk warehouse location selection. The propose hybrid method is to determine scores for all optimal locations minimizing the cost over an extended time period and the one with the highest score is finally chosen for implementation. In future the problem in this paper can be solved by Trapezoidal interval type-2 fuzzy ANP-PROMETHEE II or ELECTRA III and can also be applied in medical diagnosis, Pattern recognisition and inventory management along with other MCDM problems.

References

1. Chen, C.T.: A fuzzy approach to select the location of the distribution centre. Fuzzy Sets and systems 118, 65–73 (2001)
2. Hidaka, K.K., Okano, H.: Simulation based approach to the warehouse location problem for a large-scale real instance. In: Proceedings of 1997 Winter Simulation Conference, pp. 1214–1221 (1997)
3. Michael, L., Hentenryck, P.V.: A simple tabulation search for warehouse location. European Journal of Operational research 157(3), 576–591 (2004)
4. Sharma, R.K., Berry, V.: Developing new formulations and relaxations of single stage capacitated warehouse location problem. European Journal of Operational research 177(2), 803–812 (2007)
5. Opricovic, S., Tzeng, H.: Compromise solution of MCDM methods- A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research 156, 445–455 (2004)
6. Andrich, D.: A rating formulation for ordered response categories. Psychometrika 43, 561–573 (1978)
7. Yu, S.C., Wu, B.: Fuzzy item response model: a new approach to generate membership function to score psychological measurement. Quality and Quantity 43(3), 381–390 (2009)
8. Huang, J.H., Peng, K.H.: Fuzzy Rasch Model in TOPSIS: A new approach for generating fuzzy numbers to assess the competiveness of the tourism industries in Asian Countries. Tourism Management 33, 456–465 (2012)

Discourse Based Sentiment Analysis for Hindi Reviews

Namita Mittal, Basant Agarwal, Garvit Chouhan, Prateek Pareek, and Nitin Bania

Department of Computer Engineering,
Malaviya National Institute of Technology, Jaipur, India
nmittal@mnit.ac.in,
{thebasant,jkgarvit,prtkpareek,nitinnuts}@gmail.com

Abstract. Research on Sentiment Analysis (SA) has increased tremendously in recent times due to fast growth in Web Technologies. Hindi Language content is also growing very fast online. Sentiment classification research has been done mostly for English language. However, there has been little work in this area for Indian languages. Sentiment analysis means to extract the opinion expressed in the text about a specific topic. There is a need to analyse the Hindi language content and get insight of opinions expressed by people and various communities about a specific topic. In this paper, it is investigated that how by proper handling of negation and discourse relation may improve the performance of Hindi review sentiment analysis. Experimental results show the effectiveness of the proposed approach.

Keywords: Sentiment Analysis, HSWN, Discourse relations, negation handling, Hindi Reviews.

1 Introduction

Sentiment Analysis is a natural language processing task that deals with the findings of opinion in a piece of text with respect to a topic [9]. A large number of advertising industries and recommendation systems work on understanding the people likings and disliking's from this content. Hindi is the fourth highest speaking language in the world. The increasing user-generated content on the Internet is the motivation behind the sentiment analysis research. Majority of the existing work in this field is for English language. Very little attention has been paid in direction of sentiment analysis for Hindi Language. Information content in Hindi is important to be analysed for the use of industries.

Sentiment analysis is very difficult for Hindi language due to numerous reasons as follows. (1) Unavailability of well annotated standard corpora, therefore supervised machine learning algorithms cannot be applied. (2) Hindi is a resource scarce language; there are not efficient parser and tagger for this language. (3) Limited resources available for this language like HindiSentiWordNet (HSWN). It consists of limited numbers of adjectives and adverbs. Even, most of the words are available in inflected forms. Also, all the inflected forms of the word are not present. HSWN is created using the Hindi WordNet and English SentiWordNet (SWN). During the

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 720–725, 2013.

creation of this resource for Hindi language, it is assumed that all synonyms have the same polarity while all antonyms have the reverse polarity of a word. This assumption neglected word sense intensity in terms of polarity, however polarity intensity of their word is important in opinion mining. (4) Even, Translation dictionaries may not account for all the words because of the language variations. Same words may be used in multiple contexts and context dependent word mapping is a difficult task, error prone and requires manual efforts. Using Translation method for generating subjective lexicon, there is a high possibility of losing the contextual information and sometimes may have translation errors. In this paper, an efficient approach is proposed for identifying sentiments and opinions from user generated content in Hindi. Main objective of this paper is to investigate the influence of negation handling and discourse relations on the performance of Hindi review sentiment analysis.

This paper is organised as follows. Section 2 presents related work. Proposed approach is described in detail in Section 3. Section 4 discusses the experimental setup and results. Finally, Section 5 concludes and presents the future work.

2 Related Work

To identify the sentiment expressed in the text is difficult task for Hindi language. A lot of work has been done on sentiment analysis has been done mostly for English language [3], [5], [9], but for Hindi, sentiment analysis research in initial phase. In [2], authors created lexicon using a graph based method. They explored how the synonym and antonym relations can be exploited using simple graph traversal to generate the subjectivity lexicon. Their proposed algorithm achieved approximately 79% accuracy on classification of reviews and 70.4% agreement with human annotated. In [1], authors proposed a fallback strategy in their paper. This strategy follows three approaches: In-language Sentiment Analysis, Machine Translation and Resource Based Sentiment Analysis. The final accuracy achieved by them is 78.14 %. They developed a lexical resource, HSWN based on its English counter format. In [6], authors investigated the use of discourse and negation with the enhancement of the HSWN for Hindi reviews. In [7], authors showed that the incorporation of discourse markers in a bag-of-words model for English language improves the sentiment classification accuracy by 2 - 4%. In [4], authors proposed a method to classify Hindi reviews as positive or negative. They devised a new scoring function and test on two different approaches. They also used a combination of simple N-gram and POS-Tagged N-gram approaches.

3 Proposed Approach

Proposed approach for Sentiment Analysis of Hindi review documents works as follows. Initially, annotated dataset is created for testing of the proposed algorithm. Some basic rules are devised for negation and discourse handling which highly influence the sentiments expressed in the review. Further, HindiSentiWordNet

(HSWN) is used for the polarity values of words. Finally, overall semantic orientation of the review document is determined by aggregating the polarity values of all the words present in the document

3.1 Preparation of Annotated Dataset

Initially, 900 reviews are crawled from Hindi review websites, out of these 900 reviews, 130 reviews were rejected due to their objective nature manually. Next, for remaining 770 reviews, agreement was established on 662 reviews using Cohen's kappa. Out of these 662 total reviews, 380 were agreed as positive and 282 as negative. After that, Fleiss kappa was used for the agreement and achieved 0.8092 as kappa coefficient. This falls under the substantial agreement according to Fleiss kappa. Average size of the reviews in our dataset is 104 words.

3.2 Negation Handling

The negation operator (Example: **नही**, **न**, **नदारद** etc.) inverts the sentiment of the word following it. The usual way of handling negation in sentiment analysis is to consider a window of size n (typically 3 to 5) and reverse the polarity of all the words in the window. We reverse all the words in the window by adding (!) to every word, till either the sentence is completed or a violating expectation (or a contrast) conjunction or a delimiter is encountered. Negation on the basis of sentence structure may be applied either in forward or in backward direction. Some rules are proposed to handle negation, are discussed in following cases.

CASE 1: If a sentence has only one single negate word ("**नही**", "**नदारद**") i.e. negation is present in a simple sentence.

e.g. (1) **इस मूवी का निर्देशन अच्छा नही हैं ।** (2) **मूवी की कहानी में दम नही हैं ।**

In the above sentence, due to negation, all the words before the negation word "**नही**" would be negated and the reverse polarity of the negated words would be considered further. The above examples will be negated as

(1) **!इस !मूवी !का !निर्देशन !अच्छा नही हैं ।** (2) **!मूवी !की !कहानी !में !दम नही हैं ।**

But this negation rule may be invalid for sarcastic and special form of sentences.

e.g. **इससे बढ़िया एक्टिंग हो ही नही सकती ।**

CASE 2: If a sentence has a negation word and conjunction, and index of conjunction is more than the index of negated word, forward negation is applied. For example: (1) **फिल्म की कहानी ऐसी नही हैं कि इसे तीन घंटो तक मजे से देखा जा सके ।** (2) **बढ़िया एक्टिंग के बावजूद भी कहानी में ऐसा कुछ भी नही जो दर्शको को बांधे रख सके ।** In these sentences, negate word and the conjunction words are present and the index of conjunction is greater than the index of negate word; therefore, forward negation is applied. In above example, all the words after the conjunction will be negated .The above examples will be negated as follows. (1) **फिल्म की कहानी ऐसी नही हैं कि !इसे**

!तीन !घंटो !तक !मजे !से !देखा !जा !सके। (2) **बढ़िया एक्टिंग के बावजूद भी कहानी में ऐसा कुछ भी नही जो !दर्शको !को !बांधे !रख !सके।**

CASE 3: If a sentence have "**न**" multiple times in sub-sentences separated by commas. For example: (1) **न एक्टिंग सही हैं , न मूवी की कहानी।**

"**न**" usually occurs multiple times in this example sentence, with sub sentences separated by commas. Here for each "**न**" the negation is applied in forward direction until a delimiter is encountered. The above example will be negated as follows **न !एक्टिंग !सही ! हैं , न !मूवी !की !कहानी।।**

3.3 Discourse Relations

An essential phenomenon in natural language processing is the use of discourse relations to establish a coherent relation, linking phrases and clauses in a text. The presence of linguistic constructs like connectives, modals, and conditional can alter sentiment at the sentence level as well as the clausal or phrasal level [8]. A coherent relation reflects how different discourse segments interact. Discourse segments are non-overlapping spans of text. In this paper, Violated Expectations like **हालाकि**, **लेकिन**, **जबकि** etc. are handled. Violating expectation conjunctions oppose or refute the neighboring discourse segment. These conjunctions are categorized into the following two sub-categories: Conj_After and Conj_Infer.

3.3.1 Conj_After

It is the set of conjunctions that give more importance to the discourse segment that follows them. It means that actual segment is mostly reflected by the statement following the conjunction. So, in all the below examples, the discourse segments after the Conj_After (in bold) are given preferences and the previous sentences are dropped. For example: **लेकिन** , **मगर** , **फिर भी**, **बावजूद**

लेकिन: फिल्म की कहानी ठीक हैं, **लेकिन खराब एक्टिंग से बात बिगड़ गई।**

मगर: फिल्म इंटरवल के बाद ठीक हैं ,**मगर कुल मिलाकर वो बात नही बन पाई।**

बावजूद: **अच्छे डायरेक्शन के बावजूद भी फिल्म अपना प्रभाव नही बना पाई।**

फिर भी: **वैसे मूवी औसत हैं , फिर भी एक बार देखी जा सकती हैं।**

3.3.2 Conclusive or Inferential Conjunctions

These are the set of conjunctions, Conj_infer, that tend to draw a conclusion or inference. Hence, the discourse segment following them should be given more weight. For example: **इसीलिए** , **कुल मिलाकर**

कुल मिलाकर : कुल मिलाकर यह मूवी समय की बर्बादी हैं।

3.4 Proposed Algorithm for Sentiment Analysis of Hindi Reviews

The first step of the proposed algorithm is the pre-processing. Review documents are pre-processed by applying stemming, negation and discourse relations as discussed in previous sub-sections. After, the pre-processing, polarity values are retrieved from the HSWN. Finally, semantic orientation of the overall review document is determined by aggregating the polarity values of all the words. Proposed approach is describes in Algorithm 1.

Algorithm 1. Proposed Algorithm

Step 1: For each document in the corpus
Step 2: Apply Pre-Processing
(a) Remove the Stop Words and apply Stemming.
(b) Apply Rules (Negation and Discourse).
Step 3: For each token in the document.
Step 4: Retrieve polarity (POL) from HSWN.
Step 5: **If** (word is negated)
Then word.POL=-POL; **Else** Word.POL=POL;
Step 6: Compute the aggregate polarity of the document (doc.POL) by adding the polarities values of all the token.
Step 7: **If** (doc.POL > zero) Then label the document as positive
Else If (doc.POL<zero) Then label the document as negative
Else Classify the document as neutral.
Step 8: Return the set of Labelled Documents

4 Results and Discussions

Proposed algorithm is tested on 662 movie review dataset created by our own as described in previous section. For various experimental settings, results are reported in Table 1.

Table 1. Accuracy of various experiments

		Accuracies (In %)		
S. No	Experimental Setup	Positive	Negative	Overall
1	With only HSWN	50	51.06	50.45
2	HSWN + Negation	71.32	79.71	74.92 (+48.5%)
3	HSWN + Discourse	78.90	71.33	75.67 (+49.9%)
4	HSWN + Negation +Discourse	81.86	75.54	79.15 (+56.8%)

First of all, Semantic orientation of a document is determined by aggregating the total polarity value of all the words in the document using HSWN. Experimental results show an accuracy of 50.45%, which is very less. This accuracy is considered as baseline accuracy. The main reason for this observation was that most of the words

in our dataset were not present in the HSWN and some words are inflected forms of the available words in HSWN. Further, proposed algorithm is experimented with negation rules; it produces accuracy of 74.92% (+48.5%). Negation rules applied produces significant improvement over baseline accuracy. The main improvement due to negation was in negative reviewed documents. Further, impact of discourse relation is experimented, which gives an accuracy of 75.67% (+49.9%). Further, both negation and discourse rules are applied; it gives an accuracy of 79.15% (+56.8%).

5 Conclusion and Future Work

Opinion Mining for Hindi is an important task. In this paper, it is investigated that how the negation and discourse relations can be efficiently handled for improving the performance of sentiment analysis for Hindi reviews. Proposed approach uses the resource HSWN for the word polarity. The movie review corpus is developed in Hindi Language from the Hindi review websites. Experimental results show that proposed algorithm with negation and discourse relations significantly improves the performance for sentiment analysis. In future, the dataset can further be extended for the better and generalized results. This work can be extended to incorporate Word Sense Disambiguation (WSD) and morphological variants which could result in better accuracy for words which have dual nature. HSWN may be developed further.

References

1. Joshi, A.R., Balamurali, P.: A Fall-Back Strategy For Sentiment Analysis In Hindi: A Case Study. In: International Conference on Natural Language Processing, ICON (2010)
2. Bakliwal, A., Arora, P., Varma, V.: Hindi Subjective Lexicon: A Lexical Resource For Hindi Polarity Classification (2012)
3. Agarwal, B., Mittal, N.: Optimal Feature Selection Methods for Sentiment Analysis. In: Gelbukh, A. (ed.) CICLing 2013, Part II. LNCS, vol. 7817, pp. 13–24. Springer, Heidelberg (2013)
4. Bakliwal, P., Arora, A., Patil, V.: Towards Enhanced Opinion Classification using NLP Techniques. In: Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2011, pp. 101–107 (2011)
5. Agarwal, B., Mittal, N.: Categorical Probability Proportion Difference (CPPD): A Feature Selection Method for Sentiment Classification. In: Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology, COLING 2012, pp. 17–26 (2012)
6. Mittal, N., Agarwal, B., Chouhan, G., Pareek, P., Bania, N.: Sentiment Analysis of Hindi Review based on Negation and Discourse Relation. In: 11th Workshop on Asian Language Resources (ALR), In Conjunction with IJCNLP (in press)
7. Mukherjee, S., Bhattacharyya, P.: Sentiment Analysis in Twitter with Lightweight Discourse Analysis. In: Proceedings of the 24th International Conference on Computational Linguistics, COLING 2012 (2012)
8. Wolf, F., Gibson, E.: Representing Discourse Coherence: A Corpus-based Study. Computational Linguistics 31(2), 249–287 (2005)
9. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)

Rel-Div: Generating Diversified Query Interpretations from Semantic Relations

Ramakrishna Bairi[1], A. Ambha[2], and Ganesh Ramakrishnan[2]

[1] IITB-Monash Research Academy, IIT Bombay
[2] IIT Bombay

Abstract. Accelerated growth of the World Wide Web has resulted in an increase in appetite for searching over Internet to fulfill the information needs. Understanding user intent plays a pivotal role in determining the quality of search results and improving user satisfaction. But short, ambiguous or underspecified queries make the process of determining the concealed user intention harder. Identifying diversified but relevant interpretations for a query with an impressive accuracy, is still an active area of research. These varied interpretations originate from entities associated with the user query and relations between the entities. We address the problem of generating diverse but relevant interpretations by utilizing an Internet encyclopedia (Wikipedia) as a primary source entities and their relations. By preprocessing the encyclopedia, we build a rich repository of *Semantic Relations*, which characterize these entities and their relationships. We present algorithms to enumerate pertinent interpretations for a query based on the repository. The proposed approach uses the repository to generate candidate interpretations over which we apply graph based iterative approaches to generate diversified and relevant results. We empirically evaluate the effectiveness of our approach, with the 'Query Relevance and Understanding' dataset of TREC 2011 workshop and AMBIENT (Ambiguous Entities) dataset.

1 Introduction

The growth of internet has resulted in the proliferation of electronic documents on the World Wide Web. Every search engine, be it generic purpose or application and domain specific, serves as a portal to access these documents. User queries, in general, are short and often tend to be ambiguous and/or under-specified. In addition, a query can have multiple *concealed interpretations*. For example, *Sun* could be interpreted as "The sun as a star", "Sun Micro systems company", "Sun news paper" and so on. We believe that, in addition to these concealed interpretations, *related interpretations* are also equally important. As examples, "Solar Cells" and "Photosynthesis", could be interpretations related to this query. Out of many possible interpretations to a short query, users expect their intended answer to be present in the top few search results. This calls for the need of presenting a diversified but relevant set of results in the top K positions.

Most prior research has focused on generating diversified result [14, 16, 3, 12, 5, 17, 9, 8, 10, 2, 13, 11]. Inspired by the work GCD [5] and MMR [4], we develop a new technique for diversity ranking of interpretations based on an *interpretation graph*. As part of this technique, we propose an algorithm to learn the node and edge weights of the interpretation graph iteratively in a biconvex optimization setting.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 726–732, 2013.

2 Diversified Interpretation Generation

2.1 Our Problem

Given a large corpus $\mathcal{C}$ of m entities and a short user query q, we define a function $\mathcal{H}(q, C)$ that returns a subset of entities $\mathcal{S} \subseteq \mathcal{C}$, satisfying the query q. Let $S = \{e_1...e_n\}$. The function $\mathcal{H}(q, \mathcal{C})$ acts as a filtering function to retrieve the entities $\mathcal{S}$ that are syntactically and/or semantically related to the query q. In its simplest form, $\mathcal{H}(q, \mathcal{C})$ can just return $\mathcal{C}$ without performing any filtering, which is not generally useful. It is important to design an $\mathcal{H}(q, \mathcal{C})$ (for *e.g.*, keyword based lookup, semantics matching, *etc.*) that can help reduce the search space in a meaningful manner. We need to choose a set of K entities from $\mathcal{S}$ and we assume that to best satisfy the user intention, the K entities presented to the user should be diverse yet highly relevant to the query q. Our goal is to identify these K diverse, yet relevant results to the query q.

2.2 The Training Algorithm

We expect groups of entities in $\mathcal{S}$ to be related to each other via some semantic relations. We initially construct an entity-relation graph using $e_1...e_n$. We refer to this graph as an *Interpretation Graph*, since the entities in this graph are obtained as various interpretations of the query. While the nodes are entities from $\mathcal{S}$, each edge is a relation between the entities. A relation could be one of synonymy, hyponymy, meronymy, homonymy, *etc.*. These relations could be obtained from external catalogs such as Wikipedia, Wordnet, *etc.*

Each node in the graph is assigned a score which represents the relevance of the node to the query. We use the notation b_q to represent the column vector (of size $n \times 1$) containing all the node relevance scores. The weight on an edge represents the degree of similarity between the two nodes connected by that edge. We use the notation C_q (of size $n \times n$) to represent the matrix of edge scores reflecting similarity between pairs of nodes. Note that, each column C_q^i of the matrix C_q represents an entity e_i and the cell values in that column indicate the similarity of entity e_i with other entities. The scores in b_q are used to ensure that the subset of K interpretations are relevant to q, whereas the similarity scores in C_q are used to ensure diversity in the subset of K interpretations.

We assume that we are provided training data, consisting of queries and their correct interpretations. Our goals in training are to 1) develop a model for the node score b_q, 2) develop a model for the edge potentials C_q and 3) learn parameters of these models such that the set of K relevant yet diverse nodes obtained from the graph using b_q and C_q are consistent with the training data. Thus, implicit in our third goal is the following subproblem, which is also our query time inference problem: 4) compute a subset of K best interpretations using b_q and C_q, that represent K diverse, but relevant interpretations.

Learning Node Potentials (b_q). In order to build a learning model for b_q, it is important to define a good set of features that characterize the node's relevance to the

query. Let $N_{1..F_N}(q, \mathcal{S})$ be a set of F_N node features. Each feature $N_f(q, \mathcal{S})$ evaluates the relevance of entities in $\mathcal{S}$ to the query q and returns a vector of scores. These feature functions are problem specific and crafted carefully to bring out the relevance between query and entities. The node potential vector b_q is obtained by combining the scores returned by individual feature functions $N_f(q, \mathcal{S})$. One of the obvious choices is to use Logistic Regression [15].*i.e.* $b_q[i] = \frac{1}{1+e^{-\sum_{f=1}^{F_N} w_f N_f(q,\mathcal{S})[i]}}$. The weight vector $W^T = [w_1 ... w_{F_N}]$ is learnt through supervised training explained in Section 2.2.

Learning Edge Potentials (C_q). To learn the edge potentials, it is important to define a good set of features that measure the similarities between every pair of nodes and return similarity scores. Higher the score, more similar are the nodes. Let $C_{1..F_E}(S)$ be the set of F_E edge features that evaluate similarities between entities in $\mathcal{S}$ and each returns a $n \times n$ matrix of scores. These feature functions are problem specific and crafted carefully to bring out the similarities between the entities. The edge potential matrix C_q is obtained as $C_q = \sum_{f=1}^{F_E} \lambda_f C_f(\mathcal{S})$ where $0 \leq \lambda_f \leq 1$ and $\sum \lambda_f \geq 1 \, \forall f$. The weight vector $\lambda^T = [\lambda_1 ... \lambda_{F_E}]$ is learnt through supervised training explained in Section 2.2.

Training Feature Weights W^T, λ^T. Our technique is based on following proposition Proposition 1:

$$b_q \approx \sum_{k=1}^{K} \tilde{C}_q^{i_k} \tag{1}$$

for sufficiently large K diverse entities, where, $\tilde{C}_q$ is the matrix $C_q$with the columns scaled so that the diagonal cell values match the relevance value, *i.e.*, $\tilde{C}_q(i,i) = b_q(i)$. The values $i_1 ... i_K$ represent indices of K columns of matrix $\tilde{C}_q$. Hence, $\tilde{C}_q^{i_k}$ is the i_kth column of matrix $\tilde{C}_q$.

The intuition behind this approximated equality comes from the fact that, two similar entities should have similar relevance score with the query and we are interested in selecting K diverse entities. Let e_i be one of these K diverse entities. If the entities $e_{j_1} ... e_{j_p}$ are similar to e_i, then, $b_q[i] \approx b_q[j_1] \approx ... \approx b_q[j_p]$ and $C_q[i,i] \approx C_q[i,j_1] \approx ... \approx C_q[i,j_p] \approx 1$ and $C_q[t] \approx 0$, $t \notin j_1 ... j_p$. But, we already know that $\tilde{C}_q[i,i] = b_q[i]$. That implies, $b_q[j_1] \approx \tilde{C}_q[i,j_1]$, $b_q[j_2] \approx \tilde{C}_q[i,j_2]$, ... $b_q[j_p] \approx \tilde{C}_q[i,j_p]$. When we take the summation on all diverse K entities, the Equation 1 holds.

Based on the above observation, we present an algorithm to learn weights W^T and λ^T iteratively in a supervised learning setup. The training data is provided in a vector r_q (of size $n \times 1$) such that $r_q[i] = 1$ if the entity e_i is relevant to the query (and one of diverse entities), otherwise, $r_q[i] = 0$. Note that, the quantity $\tilde{C}_q r_q$ represents the sum of K columns (assuming K number of 1s in r_q) and is the RHS of Equation 1.

Our training objective is to learn λ^T and W^T such that Equation 1 holds. Formally, the problem being solved is:

$$\underset{\lambda_1 ... \lambda_{F_E}, w_1 ... w_{F_N}}{argmin} \; D\left(\frac{1}{1+e^{-\sum_g w_g N_g}}, \sum_f \lambda_f \tilde{C}_f r_q \right) \tag{2}$$

where $D(x, y)$ is a distance measure between x and y. (for e.g., KL Divergence, Euclidean, etc.); $\tilde{C}_f$ is the normalized C_f as in Proposition 1.

We learn the weights W^T and λ^T iteratively using two steps outlined in Equation 3 and Equation 4, each of them convex in the respective optimization variables, resulting in binconvex optimization.

div-step: Learn $\lambda_1^{(t)}, \lambda_2^{(t)}, ...$ holding $w_1^{(t-1)}, w_2^{(t-1)}, ...$ constant, by solving:

$$\underset{\lambda_1,\lambda_2,...}{argmin}\ D\left(\frac{1}{1+e^{-\sum_g w_g^{(t-1)} N_g}}, \sum_f \lambda_f^{(t)} \tilde{C}_f r_q\right) \tag{3}$$

rel-step: Learn $w_1^{(t)}, w_2^{(t)}, ...$ holding $\lambda_1^{(t-1)}, \lambda_2^{(t-1)}, ...$ constant, by solving:

$$\underset{w_1,w_2,...}{argmin}\ D\left(\frac{1}{1+e^{-\sum_g w_g^{(t)} N_g}}, \sum_f \lambda_f^{(t-1)} \tilde{C}_f r_q\right) \tag{4}$$

In *div-step*, we learn λ^T by holding W^T fixed and honoring Equation 1. In *rel-step*, we learn W^T by holding λ^T fixed. r_q is provided by the user as part of training data.

We learn node and edge feature weights iteratively by recognizing and assigning weights to prominent node and edge features that satisfy queries of different types. Having all statistically driven computation of weights for edge features can minimize the side effect of poor node features and likewise computing weights for node features can decrease the consequences of poor edge features.

Algorithm 1 outlines the training procedure. I_q^+, I_q^- are the set of relevant and irrelevant entities for each query q in the ground truth used for training.

Inference. For a new user query q, inference problem is to choose K diversified results. Using $\mathcal{H}(q, \mathcal{C})$ we reduce the search space drastically and get the set $\mathcal{S}$. Otherwise, we need to run our inference on entire set $\mathcal{C}$, which is very expensive. We then compute the node and edge feature matrices for all defined node and edge features. These individual feature matrices are then combined (using λ^T and W^T) to obtain vector b_q and matrix C_q. Based on Proposition 1, our inference objective is to choose K columns from the matrix $\tilde{C}_q$ such that their sum is as close as possible to b_q. Formally, the problem being solved is:

$$\underset{i_1...i_K}{argmin}\ D\left(b_q, \sum_{k=1}^{K} \tilde{C}_q^{i_k}\right) \tag{5}$$

where $i_1...i_K$ are indices of K columns of matrix $\tilde{C}_q$.

Finding out exact solution (i.e. $i_1...i_K$ columns) to the above optimization problem turns out to be computationally infeasible. Algorithm 2 describes a greedy inference procedure. At each step we pick one column from $\tilde{C}_q$ that minimizes the distance in Equation 5 most. However, we also ensure that the picked column is most diverse from the already selected columns in the previous steps.

Algorithm 1. Training

Input: Set of training data instances $\left\{q, I_q^+, I_q^-, N_f, C_f, r_q\right\}$
Output: W^T and λ^T
initialize variables W^T and λ^T
learn initial W^T using Logistic Regression
▷ uses $\left\{q, I_q^+, I_q^- N_f\right\}$
▷ $\tilde{C}_q, \tilde{C}_f$ used below are normalized C_q, C_f as in Proposition 1
while not converged($\mid b_q - \tilde{C}_q r_q \mid$) **do**
    b_q= compute relevance matrix using W^T and I_q^+
    find λ^T so that $D\left(b_q, \sum_f \lambda_f \tilde{C}_f r_q\right)$ is minimized
    ▷ W^T is fixed
    $p_q = \sum_f \lambda_f \tilde{C}_f r_q$
    find W^T so that $D\left(\frac{1}{1+e^{-\sum_f w_f N_f}}, p_q\right)$ is minimized
    ▷ λ^T is fixed
end while
return $\left(W^T, \lambda^T\right)$

Algorithm 2. Inference

Input: $q, \mathcal{C}, \lambda^T, W^T, N_f, C_f$
Output: K diverse interpretations
Generate $\mathcal{S} = \mathcal{H}(q, C)$ and build a graph using entities in $\mathcal{S} = \{e_1, .., e_n\}$
Compute b_q using W^T and node features $N_{1..F_E}(q, \mathcal{S})$
Compute $C_q = \sum_f \lambda_f C_f(\mathcal{S})$ and normalize as in Proposition 1
$R = \{$ï¿œ$\}$ ▷ set of selected indices
$Q = \{i_1, .., i_n\}$ ▷ indices to select
for $i = 1$ to K **do**

$$\underset{c_k \in Q/R}{argmin} \left\{ D\left(b_q, \sum_{r \in R \cup \{c_k\}} \left(C_q^r\right)\right) \times \left(1 - \frac{1}{Z} min\left(D\left(C_q^{R_1}, C_q^{c_k}\right), ..., D\left(C_q^{R_{|R|}}, C_q^{c_k}\right)\right)\right) \right\}$$

▷ (query match) × (dissimilar to selected), z is normalizer
    $R = R \cup \{c_k\}$
end for
return K interpretations representing K columns $R_1, ..., R_{|R|}$

3 Experimental Evaluation

We apply our Rel-Div technique to generate interpretations to a short and/or ambiguous user query (*e.g. Beagle, Sony Camera*, etc) using Wikipedia. We do not support queries which are highly rich in semantics like *Who invented music* or very specific in nature like *DB2 error code 1064*. We used various signals from Wikipeida (such as title, hyperlink structure, infoboxes, category structure, etc) to define a bunch of node features and edge features. We evaluated our technique using the QRU dataset of SIGIR 2011 contains 100 TREC queries and the AMBIENT dataset which contains 40 one word queries.

3.1 Evaluation Methodology

The relevance of any interpretation to the query is measured using precision at different positions and the diversity is estimated using NDCG-IA [1]. For Recall, since it is practically not possible to manually inspect all Wikipedia entities for relevance to a query, we based our recall on the candidate interpretations generated.

In our experiments, we consider a couple of other approaches to diversification, which have been reported in literature, though used in other problem settings. These include variants of GCD [5] and Affinity Propagation [6, 7].

M-Div *:* Uses page rank matrix M as in GCD instead of the C_q matrix.

M-Div-NI: Similar to M-Div, but node and edge weights are learnt independently (as in GCD).

AFP*:Exemplar* nodes of Affinity propagation are taken as interpretations.

3.2 Comparisons of Approaches

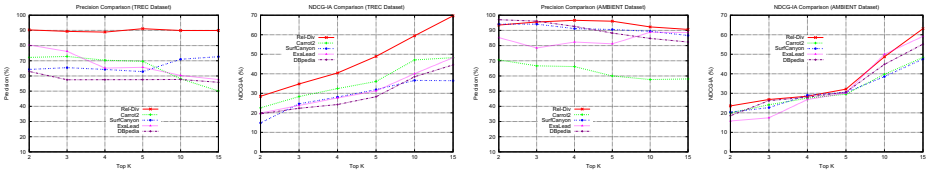
While experimenting with our proposed approach, we found best performance when D in div-step was chosen to be KL-divergence and D in rel-step was chosen as the Euclidean distance. In Table 1, we compare the proposed diversification algorithm against M-Div, M-Div-NI and AFP on precision, recall and NDCG-IA measures.

We observed that our Ranking algorithm Rel-Div performs at par with (and sometimes even better than) M-Div and M-Div-NI. However, one of the major advantage of our method compared to M-Div and M-Div-NI is that, we need not calculate the inverse of C_q matrix, which is a computationally intensive process for a large dimension matrices.

We also compare the diversity in interpretations using our approach against those from four other search engines: Carrot2, SurfCanyon, Exalead and DBPedia. Figure 1 confirms that our approach produces better diversity in the interpretations.

Table 1. Results of different approaches

		Precision(%)			Recall(%)			NDCG-IA(%)		
		@5	@10	@10	@5	@10	@10	@5	@10	@10
TREC	Rel-Div	**91.13**	**89.93**	**89.83**	**7.02**	**13.85**	**20.4**	48.9	59.47	**69.7**
	M-Div	89.87	84.27	84.32	6.74	12.71	18.88	**49.71**	**62.68**	67.39
	M-Div-NI	83.75	80	80	6.83	12.81	19.35	42.48	60.88	66.52
	AFP	78.3	76.9	80.7	6.3	12.4	18.1	34.2	38.8	47.6
AMBIENT	Rel-Div	96.05	92.3	90.67	7.33	**14.57**	**21.61**	**32.12**	**48.72**	**63.1**
	M-Div	96.15	**94.15**	**93.56**	**7.43**	14.37	21.61	32.41	47.49	58.09
	M-Div-NI	**96.2**	93.58	93.19	7.33	13.87	21.11	22.93	43.59	55.84
	AFP	88.4	90.9	92.3	6.9	13.6	21.47	32.09	45.9	55.1



(a) Precision : TREC (b) NDCG-IA : TREC (c) precision : AMB (d) NDCG-IA : AMB

Fig. 1. Comparison with external systems

Examples of interetations generated from our system for many short queries can be found at `http://qassist.cse.iitb.ac.in/facetedsearcher/examples.pdf`

4 Conclusion

We presented techniques for generating the top K interpretations to a user query using some internet encyclopedia, (in particular, Wikipedia). Our approach caters to two needs of the user, *viz.*, that all the interpretations are relevant and that they are as diverse as possible. We addressed this using a bunch of node features and edge features based

on semantic relations and learn these feature weights together iteratively. Our experimental evaluations, comparison to existing techniques and publicly accessible systems show that our approach performs well on both the fronts. We believe technique can be improved for better handling of multiword queries by adopting deep NLP parsing techniques, which will form part of our future work.

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying search results. In: WSDM 2009, pp. 5–14. ACM, New York (2009)
2. Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E.J., Sznajder, B., Yogev, S.: Beyond basic faceted search. In: WSDM, pp. 33–44 (2008)
3. Brandt, C., Joachims, T., Yue, Y., Bank, J.: Dynamic ranked retrieval. In: WSDM, pp. 247–256 (2011)
4. Carbonell, J., Goldstein, J.: The use of mmr, diversity-based reranking for reordering documents and producing summaries. In: Research and Development in Information Retrieval, pp. 335–336 (1998)
5. Dubey, A., Chakrabarti, S., Bhattacharyya, C.: Diversity in ranking via resistive graph centers. In: KDD 2011, pp. 78–86. ACM, New York (2011)
6. Frey, B., Dueck, D.: Mixture modeling by affinity propagation. In: Advances in Neural Information Processing Systems 18, pp. 379–386. MIT Press, Cambridge (2006)
7. Frey, B.J., Dueck, D.: Clustering by passing messages between data points. Science 315 (2007)
8. Hahn, R., Bizer, C., Sahnwaldt, C., Herta, C., Robinson, S., Bürgle, M., Düwiger, H., Scheel, U.: Faceted wikipedia search. In: Abramowicz, W., Tolksdorf, R. (eds.) BIS 2010. LNBIP, vol. 47, pp. 1–11. Springer, Heidelberg (2010)
9. Hearst, M.A.: Clustering versus faceted categories for information exploration. Commun. ACM 49(4), 59–61 (2006)
10. Li, C., Yan, N., Roy, S.B., Lisham, L., Das, G.: Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia. In: WWW, pp. 651–660 (2010)
11. Ma, H., Lyu, M.R., King, I.: Diversifying query suggestion results. In: AAAI (2010)
12. Raman, K., Joachims, T., Shivaswamy, P.: Structured learning of two-level dynamic rankings. In: CIKM (2011)
13. Shen, Y., Yan, J., Yan, S., Ji, L., Liu, N., Chen, Z.: Sparse hidden-dynamics conditional random fields for user intent understanding. In: WWW 2011, pp. 7–16. ACM, New York (2011)
14. Swaminathan, A., Mathew, C.V., Kirovski, D.: Essential pages. Web Intelligence, 173–182 (2009)
15. Yan, L., Dodier, R.H., Mozer, M., Wolniewicz, R.H.: Optimizing classifier performance via an approximation to the wilcoxon-mann-whitney statistic. In: ICML, pp. 848–855 (2003)
16. Yue, Y., Joachims, T.: Predicting diverse subsets using structural SVMs. In: ICML, pp. 271–278 (2008)
17. Zhai, C., Cohen, W.W., Lafferty, J.: Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In: SIGIR, pp. 10–17 (2003)

Automatic Generation of Multiple Choice Questions Using Wikipedia

Arjun Singh Bhatia, Manas Kirti, and Sujan Kumar Saha

Department of Computer Science and Engineering
Birla Institute of Technology, Mesra, Ranchi, India - 835215
{arjunbhatia2304,manasmikku21,sujan.kr.saha}@gmail.com

Abstract. In this paper we present a system for automatic generation of multiple choice test items using Wikipedia. Here we propose a methodology for potential sentence selection with the help of existing test items in the web. The sentences are selected using a set of pattern extracted from the existing questions. We also propose a novel technique for generating named entity distractors. For generating quality named entity distractors we extract certain additional attribute values on the key from the web and search the Wikipedia for the entities having similar attribute values. We run our experiments in sports domain. The generated questions and distractors are evaluated by a set of human evaluators using a set of parameters. The evaluation results demonstrate that the system is reasonably accurate.

1 Introduction

Multiple choice question (MCQ) is a very popular form of assessment in which respondents are asked to select the best possible answer out of a set of choices. A MCQ is composed of three elements: stem, target word and distractors. The *stem* (also known as *item*) is the sentence from which the question is formed, *target word* (also named as *key*) is the correct answer of the question and *distractors* are the set of wrong answers.

Development of automatic MCQ generator has become a popular research problem in the last few years. In the literature we observe, generally automatic MCQ systems have been followed three major steps: selection of sentences (or stem), selection of target word and generation of distractors. Mitkov and Ha (2003) and Mitkov et al. (2006) proposed a NLP-based methodology for generating MCQ semi-automatically from an electronic text, a textbook on linguistics. They used several NLP techniques, natural language corpora and WordNet. They have used various post editing phases for betterment of the system. Aldabe et al. (2006) and Aldabe and Maritxalar (2010) developed another system to generate MCQ in Basque language in the science domain. They divided the task into six phases: selection of text (based on level of the learners and the length of the texts), marking of blanks (done manually), generation of distractors, selection of distractors, evaluation with learners and item analysis. Papasalouros et al. (2008) proposed an ontology based approach for development of an automatic

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 733–738, 2013.

MCQ system. They used the structure of an ontology - the concepts, instances and the relationship or properties that relates the concepts or instances - to generate the MCQs. First they formed sentences from the ontology structure and then they found distractors from the ontology. Agarwal and Mannem (2011) presented a system for generating gap-fill questions, a problem similar to MCQ, from a biology text book. For sentence selection they used a number of features like, is it first sentence, contains token that occurs in the title, position of the sentence in the document, whether it contains abbreviation or superlatives, length, number of nouns and pronouns etc. Similarly for key selection and distractor generation they used sets of relevant features.

In this paper we present a novel technique for MCQ generation. Our system does not require any ontology or WordNet; rather it is using the web, specially the Wikipedia, as the source of information. Therefore it can be easily transferred to any domain or language. Currently we run the experiments in sports domain. We have divided the task into several subtasks which are summarized below and discussed in the remaining sections of the paper.

- Find some available MCQs on the domain of interest from the web. Form sentences from these collected MCQs. This set is called as 'reference set' in our experiments.
- Search for potential sentences from the Wikipedia. For that we extract patterns from the reference set sentences and find sentences that are containing these patterns.
- Select *key* from the potential sentences. Most of the keys are *named entity* (NE) in this domain.
- Form question from the sentences.
- Generate *distractors* using Wikipedia. A technique is proposed for generating interesting and relevant NE distractors.

2 Sentence Selection and Question Generation

We have already mentioned that we have generated the MCQs from Wikipedia text. A typical Wikipedia page contains hundreds of sentences and thousands of words. For example, the Wikipedia page on *Sachin Tendulkar* contains more than 650 sentences and about 11000 words. MCQ can not be formed from all the sentences. Therefore first of all we have to select the potential MCQ sentences. Again a selected potential sentence contains several words; we need to select the word (or word n-gram) which can become the key. Next we form the question. These phases of the task are described below.

Reference Set Generation: For selecting the potential sentences we have taken the help of existing MCQs on this domain. MCQ based assessment is quite popular and multiple choice question papers are often generated (manually) for assessment purpose. A number of such MCQ papers are available in the web. We collect such available MCQs from different sources.

Next we form sentences from the MCQs. MCQs contain the stem and a few options. For sentence generation we replace the 'Wh phrases' (who, where, which,

when, which of the following etc.) in the stem by the first option. The first option might not be the correct answer of the question, but our intention is to sentences that are syntactically correct. We refer this set of generated sentences as 'reference set'.

Sentence Extraction Using Patterns: We extract a set of context patterns from the reference set sentences for finding potential sentences. For extracting the patterns we first run stemmer (for getting the root words for the inflected words) and the Stanford NER system[1]. The identified NEs are replaced by variables (for example, *PER* for person NEs, *LOC* for location entities). Then find the most frequent n-grams (where n is taken upto seven and minimum length is three) that are not occurring in general domain pages. These are the patterns. Some example patterns are, *PER be the captain of LOC*, *PER be the middle name of PER*, *PER word* the man of the tournament*, (word* can be replaced by zero or more words) *maximum number of one day ducks*, *fastest one day (or, word*) century*, *most successful bowler of the tournament*, *opening ceremony was held in* etc.

Key Identification: Key identification is the next phase where we select the word (or n-gram) that has the potential to become the key. We have identified the potential sentences using the patterns. A pattern is likely to extract sentences containing a particular type of entities. For example, "fastest one day century" pattern should retrieve sentences containing the name of the cricketer having the fastest century. Therefore the key for this pattern should be the person name in the retrieved sentence. Similarly, "the man of the tournament" pattern will extract sentences having the name of the player who got the man of the tournament in a particular tournament. The key for the pattern should be the person name. The pattern "opening ceremony was held in" is expected to retrieve the location (city name or ground name) where the opening ceremony of a tournament was held; therefore the corresponding key will be the location entity. For each of the extracted patters we identify the entity type which is having the potential to become the key. The sentences are tagged using the NER system and the corresponding entity is selected as the key.

Question Formation: For question formation also we have taken the help of the patterns. The patterns give us the information regarding the type of the key. Depending on the key category we replace the key by a proper *wh-word*. The parse structure of the sentence is also consulted to bring the wh-word in the beginning of the question. If the category is person then the wh-word is *who*; similarly, for location *where*, for date *when*, for number *how many* etc.

3 Distractor Generation

Distractor is a concept semantically close to the keyword. Quality of the distractors plays an important role in MCQ. In this article we have focussed in the

[1] `http://nlp.stanford.edu/software/CRF-NER.shtml`

sports domain where most of the distractors are named entity. Here selection of distractors is more challenging. For example, consider the question "Who was the Indian wicketkeeper in the final match of Cricket World Cup 2011?". And if the generated options are, Ricky Ponting, Adam Gilchrist, Chris Gayle and Mahendra Singh Dhoni then quality of the MCQ is somehow degraded. Because among the options Mahendra Singh Dhoni is only from India and is becoming the obvious choice. For selecting a set of good and interesting distractors we use certain additional information extracted from the web regarding the particular stem.

For distractor generation WordNet, domain ontologies or related knowledge base for finding similar or related words (synonyms, hypernyms, hyponyms, antonyms etc.) are often used in the literature. Mitkov et al. (2009) experimented on several similarity methods for distractor generation: collocation pattern, four different methods of WordNet based semantic similarity, distributional similarity and phonetic similarity. Brown et al. (2005) also used WordNet for finding distractors. Correia et al. (2010) studied random, graphemic and phonetic approaches for automatic distractor generation. Aldabe and Maritxalar (2010) used Latent Semantic Analysis (LSA), dictionary and WordNet to generate distractors. Papasalouros et al. (2008) used the ontology to generate distractors. They have chosen distracters as the related instances or classes having similar properties with the key. Agarwal and Mannem (2011) generated distractors using three features namely, contextual similarity, sentence similarity and term frequency, along with the parts-of-speech information.

3.1 Our Approach for Distractor Generation

In this cricket domain the major category of key (or distractors) are: person name (cricketer, bowler, batsman, wicketkeeper, captain, man-of-the-match etc.), name of team (country name, club or franchise name etc.), location name (name of cricket ground, name of city etc.), date (year, match date etc.), numerical (number of matches, batting average, number of wicket, highest score etc.) tournament name (cup, trophy, championship etc.) etc. We develop a framework for handling all these categories. Basically all the categories are handled using similar approach, only the *attribute set* representing the category differs. In the following we have discussed in detail the strategy for generating person name distractors.

For generating the distractors for a person name we find a set of additional attributes for the key from the Wikipedia. The attribute set for the cricketer contains: date of birth (or born), role (bowler, batsman etc.), batting style, bowling style, national side, other team, last match played (ODI/Test/T20), is captain (binary value, current or former captain). These information are extracted from the tabular or structured data (information box on the title) present in the right hand side top of an Wikipedia page or the first sentence of the content.

Next we search the Wikipedia for a list of related candidates from the same category. For example, for cricketer key we feed a search query "list of <national side> cricketers"; if the *is-captain* attribute value is one then the query is "List

Table 1. Performance of the MCQ generator

Evaluator	Domain	Key	Question formation	Info	Distractors	Close
Evaluator 1	100	76	70	20	81	80
Evaluator 2	100	84	72	27	85	84
Evaluator 3	100	81	74	18	96	85
Evaluator 4	100	77	68	20	89	88
Evaluator 5	100	79	73	21	89	92
Average	100	79.4	72.6	21.2	88	85.8

of <national side> national cricket captains". Wikipedia search engine retrieves page containing the lists given in tabular form. The table consists of a set of fields. For example, the list of cricketer contains the fields like, span, number of matches played, batting, bowling, and fielding summary. From the table we extract a set of cricketers having attribute values similar as the key. This set acts as *candidate distractors*.

Then we check whether there is any special feature in the stem; for example, man-of-the-match of a particular match, highest wicket taker of a tournament. If any such is there then we extract the list of players played the match or tournament, most run getter or most wicket taker statistics of the tournament etc. These entities get higher preference to become the distractors. Otherwise randomly pick the required distractors from the candidate set.

4 System Evaluation and Discussion

In order to evaluate the quality of the MCQs generated by the system, we define a set of parameters. Item analysis has been used for MCQ evaluation in a few articles which consists of three parameters (Gronlund, 1982): difficulty of the item, discriminating power and usefulness of each distractor. In this paper we have extracted the items from Wikipedia and majority of the distractors are NE, therefore here we have used a extended set parameters. These are, whether the question is relevant to the domain of interest (referred in table as *domain*), the key is chosen properly, question is formed properly, question is over-informative or under-informative (*info*), whether the distractors are related with the key (*distractor*), at least one distractor is close to key (*close*). *Question is over-informative or under-informative* is a special parameter required when the sentences are extracted from the web using patterns. From the Wikipedia page on ICC World Cup 2011 our system form a MCQ as "who was the man of the tournament?" But for which tournament? This information was not there in that particular sentence which is not required to mention repeatedly as the whole article is on World Cup 2011. But the question is under-informative and cannot be answered. Closeness it is another special parameter required in this task. For example, in question related to man-of-the-match at least one of the distractors should have played that match and also performed well.

The quality of the generated MCQs is evaluated by five human evaluators using the aforementioned parameters. We have generated 100 MCQs by the

system and given to the human evaluators. In Table 1 we present the accuracy of the system. From the table we observe that the system is generating good quality distractors, the average accuracy is 88%. The system also selected the key with high accuracy (about 80%). But in question generation the accuracy is 72.6%. Around 28% questions are not formed properly. We have analyzed those questions - the errors occurred mainly because of lengthy, compound or complex sentences. Also 21.2% questions are over-informative or under-informative.

5 Conclusion

In this paper we have presented a system for automatic MCQ generation in sports domain. The system is using web information, specially Wikipedia, for generating questions and distractors. We have attempted to find the potential MCQ sentences with the help of existing questions of this domain. Our system generates named entity distractors of good quality. Finally the system is tested by a set of human evaluators using a proposed set of parameters. Evaluation results demonstrates that the system produces the MCQs and distractors with reasonable accuracy.

References

Agarwal, M., Mannem, P.: Automatic Gap-fill Question Generation from Text Books. In: Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications, pp. 56–64 (2011)

Aldabe, I., de Lacalle, M.L., Maritxalar, M., Martinez, E., Uria, L.: ArikIturri: An Automatic Question Generator Based on Corpora and NLP Techniques. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 584–594. Springer, Heidelberg (2006)

Aldabe, I., Maritxalar, M.: Automatic Distractor Generation for Domain Specific Texts. In: Loftsson, H., Rögnvaldsson, E., Helgadóttir, S. (eds.) IceTAL 2010. LNCS (LNAI), vol. 6233, pp. 27–38. Springer, Heidelberg (2010)

Brown, J.C., Frishkoff, G.A., Eskenazi, M.: Automatic question generation for vocabulary assessment. In: Proceedings of HLT/EMNLP, pp. 819–826 (2005)

Correia, R., Baptista, J., Mamede, N., Trancoso, I., Eskenazi, M.: Automatic Generation of Cloze Question Distractors. In: Second Language Studies: Acquisition, Learning, Education and Technology (2010)

Gronlund, N.: Constructing achievement tests. Prentice-Hall Inc., New York (1982)

Mitkov, R., Ha, L.A.: Computer-aided generation of multiple-choice tests. In: Proceedings of the HLT/NAACL 2003 Workshop on Building educational applications using Natural Language Processing, pp. 17–22 (2003)

Mitkov, R., An, L.A., Karamanis, N.: A computer-aided environment for generating multiple-choice test items. Journal of Natural Language Engineering 12(2), 177–194 (2006)

Mitkov, R., Ha, L.A., Varga, A., Rello, L.: Semantic similarity of distractors in multiple-choice tests: extrinsic evaluation. In: Proceedings of the EACL 2009 Workshop on GEometical Models of Natural Language Semantics, pp. 49–56 (2009)

Papasalouros, A., Kanaris, K., KotisAutomatic, K.: Generation of multiple-choice questions from domain ontologies. IADIS e-Learning (2008)

Link Prediction Using Power Law Clique Distribution and Common Edges Distribution

Srinivas Virinchi* and Pabitra Mitra

Dept of Computer Science and Engineering,
Indian Institute of Technology
Kharagpur-721302, India
{virinchimnm,pabitra}@cse.iitkgp.ernet.in

Abstract. Link Prediction is an interesting problem and is concerned with predicting important edges in a social network based on the current link structure. This prediction is based on the similarity between the two nodes; similarity is captured typically using some function of the degree of the common neighbors of the two nodes. The well-known power law degree distribution is helpful in designing relevant functions used in computing similarity functions. We show that cliques of nodes in the graph also follow a power law distribution in terms of their size. We call this power law clique distribution. It prompts us to consider small size cliques in computing similarity. We specifically use cliques of size three in an appropriately weighted form to compute the similarity. Cliques of size three correspond to common edges. By using the proposed similarity functions, we show experimentally an improvement in performance in terms of classification accuracy over the state-of-the-art local similarity functions using benchmark datasets.

Keywords: Common Neighbors, Resource Allocation Index, Cliques of size two and three, Power law.

1 Introduction

In this paper we deal with the link prediction problem. Link prediction problem deals with computing similarity between two nodes, which are not connected, using local neighborhood or globally. In [1] and [2] surveys of various local and global approaches in computing the similarity are presented. They show that similarity measures like Common Neighbors and Resource Allocation Index are comparable to the state-of-the-art similarity measures in terms of classification accuracy. Here, we deal with some of the state-of-the-art local similarity measures. Another successful similarity measure is Resource Allocation Index. According to this measure, the probability of a link formation between two unconnected nodes considers the common neighbors into account and takes the contribution from each of them as inversely proportional to their degree. The

* Corresponding author.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 739–744, 2013.

authors in [4] designed this metric which was motivated by the resource allocation dynamics where the resource is transmitted through the common neighbor based on its capacity which is reflected by its degree. Most of these metrics use some function of the common neighbors and weigh their contribution accordingly. Similarity computation between two nodes based on communities is proposed in [5]; here they consider both common nodes and common edges from different communities to which both the nodes belong. They give equal importance to all the common edges and communities. They show improvement in classification accuracy. However, they do not exploit the power law degree distribution.

Typically social networks follow the power law degree distribution [3]. We exploit the power law degree distribution to show experimentally that there is a power law associated with the sizes of cliques. This permits us to extend the existing similarity measures to take into account common edge information; we ignore the contribution of edges between a pair of high degree nodes. Specifically, we have compared the performance of the modified algorithms with common neighbors and resource allocation index. We consider the role of common edges between nodes where one of them is a low degree node. The modified algorithms have resulted in an improved performance in terms of classification accuracy on several benchmark datasets on collaboration networks.

Our specific contributions in this paper are:

1. We establish power law distribution of clique sizes based on power law degree distribution.
2. We devise schemes to ignore the contribution of common edges between high degree nodes appropriately based on power law distribution of clique sizes.

This approach can be integrated with any local similarity function to improve classification accuracy. We have shown experimentally that the proposed schemes are superior to their counterparts, viz. similarity computation based on common neighbors and resource allocation index. The rest of the paper is organized as follows: section 2 introduces the background and some of the standard state-of-the-art similarity functions and the modifications that we suggest. In section 3, we describe the experimental results. We conclude in section 4.

2 Background and Our Approach

A social network is a graph $G = (V, E)$ where V is the set of vertices and E is the set of edges in the graph [3]. Now, let us consider that at some future instance t the graph after addition of some edges has become $G_t = (V, E_t)$ where E_t is the set of edges. Link Prediction problem deals with the prediction of edges from the set of edges $E_t - E$ accurately.

For solving the link prediction problem, we make use of cliques. A clique in G is any completely connected subgraph of G. We conducted experiments on the benchmark datasets to check if they follow a power law distribution. These are shown in figure 1. Note the curve fitted in figure 1, which follows $p_k = Ck^{-\alpha}$ where $C = 3$ and $\alpha = 2$ where k is the size of cliques which ranges from 3

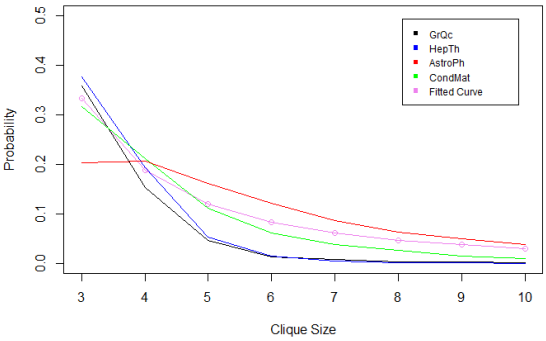


Fig. 1. Power Law Distribution of clique sizes between 3 and 10

to 10 and p_k is the probability of observing a clique of size k; there are a large number of small size cliques and a small number of large size cliques which can be approximated to a power law distribution [4] of clique sizes as shown in figure 1. For the sake of uniformity across all the datasets we considered cliques up to size 10. In figure 2, we show two cliques of size 3; ABD and ACD sharing a common edge AD. In our method we use cliques of size 3 to find out the common edges between two unconnected nodes. By common edge between two nodes B and C, we refer to the edge (A, D) that is common to both the cliques ACD and ABD of size 3 each as shown in figure 2. Futher, notice that there can be atmost one edge common to two cliques of size 3. Note that we are considering cliques of size 3 to find the common edges and we ignore large size cliques as they are infrequent according to the power law distribution of cliques shown in figure 1.

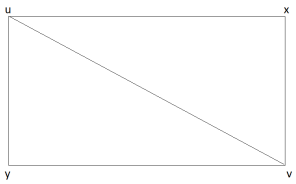


Fig. 2. Cliques ACD and ABD sharing common edge AD

We make use of the following similarity functions and modify them suitably by weighing the contribution from the common edges as explained below.

Common Neighbors(CN) : Score between nodes x and y is calculated as the number of common neighbors between x and y. Here, $N(x)$ is the set of nodes adjacent to node x.

$$CN(x, y) =\mid N(x) \cap N(y) \mid$$

Resource Allocation Index (RA): Score between nodes x and y is calculated as the sum of inverse of the degree of each of the common neighbors z between x and y.

$$RA(x,y) = \sum_{z \in N(x) \cap N(y)} \frac{1}{degree(z)}.$$

Now we present our modifications to the above metrics. Let us call them **CN Extended (CNE)** and **RA Extended (RAE)** which stand for the modified similarity functions for CN and RA respectively.

$$CNE(x,y) = CN(x,y) + CE(x,y) \;\; where$$

$$CE(x,y) = \begin{cases} 0 & \text{if } degree(x) > T \text{ and } degree(y) > T \\ \sum_{(u,v)} 2 & \text{else} \end{cases}$$

$$RAE(x,y) = RA(x,y) + RE(x,y) \;\; where$$

$$RE(x,y) = \begin{cases} 0 & \text{if } degree(x) > T \text{ and } degree(y) > T \\ \sum_{(u,v)} 2/(\text{degree(u)}+\text{degree(v)}) & \text{else} \end{cases}$$

Here, (u, v) refers to the common edge shared by a cliques xuv and yuv as shown in figure 2. A threshold value T is used to ignore the contribution of high degree nodes to CE and RE as shown in the above equations. In order to ignore the higher degree nodes in the graph we set the value of T to $0.5 * max\{degree(x)\}(\forall x \in V)$; doing this we ignore contributions of a very small number of nodes as the datasets follow a power law degree distribution [3].

From our approach it is clear that we are ignoring the contribution of common edges between two high degree nodes (in both RE and CE) and giving more importance to common edges between two low degree nodes (in RE). Here, we choose two as a factor in the expression for CE and RE as each common edge has two end vertices u and v whose contribution are weighed in a manner similar to to that of CN and RA respectively to account for the common edges. In CE, we give equal score to all the edges whereas in RE we give more importance to the edge if the end vertices u and v are of lower degree which varies as $2/(degree(u) + degree(v))$. We also ensure that the CE and RE score that is computed do not dominate the CN and RA scores respectively. The CE and RE score is zero if the end vertices u and v both have degrees greater than the threshold T.

3 Experimental Results

Standard collaboration benchmark datasets from [6] are used for experimentation which are undirected graphs. The nodes represent the authors and edges

represent collaborations. Each dataset is preprocessed by removing self-loops in the corresponding graph and is tabulated in table 1.

We setup the experiment in a manner similar to the one explained in [2] and [5]. We randomly partition the edges of the preprocessed graph into five distinct edge sets. For each fold, we use four of the edge sets as the training graph and predict the edges in the fifth part (test data). We repeat this five times taking a different part to be the test data each time; this is called *5-fold cross-validation*. For each fold, we compute the classification accuracy as the percentage of edges that are correctly predicted to have an edge in the test data. We take the average over 5 folds and list the classification accuracy on the datasets by using the standard and the modified metrics and list them in table 1.

Table 1. Percentage Classification Accuracy

Dataset	**| V |**	**| E |**	**T**	**CN**	**CNE**	**RA**	**RAE**
GrQc	5241	28968	65	98.4	**99.3**	98.4	**99.5**
HepTh	9875	51946	53	78.1	**81.2**	92.2	**94.36**
CondMat	23133	186878	45	81.61	**91.1**	96.76	**97.14**
AstroPh	18771	396100	322	98.81	**99.05**	99.45	**99.55**

Table 2. Percentage Classification Accuracy on HepTh Dataset for 5-folds

Sl. No	**Predicting top 5% edges**				**Predicting top 10% edges**				**Predicting top 15% edges**			
	CN	**CNE**	**RA**	**RAE**	**CN**	**CNE**	**RA**	**RAE**	**CN**	**CNE**	**RA**	**RAE**
1	99.2	98.4	95.6	100	80.2	82.2	93.4	94.2	66.8	70.4	84.93	84.93
2	98.8	94.8	95.2	100	80.8	81.4	90.6	95	66.4	73.6	86.26	84.8
3	96.4	97.6	96.8	99.6	72.2	80.6	92	93.6	61.46	69.3	85.2	87.73
4	97.6	98.4	96.4	100	76.0	81.2	91.8	94	65.33	70.67	85.7	86.4
5	99.6	97.6	98.8	99.2	81.4	80.4	93.2	95	67.86	70.13	86.5	87.46
Total	98.3	**97.3**	96.5	**99.76**	78.12	**81.16**	92.2	**94.36**	65.57	**70.82**	85.7	**86.26**

From table 1 we observe that RA outperforms CN which is consistent with results shown in [2]. Note that CNE and RAE perform better than CN and RA respectively as shown in boldface; RAE is the best. There is an increase in classification accuracy up to 9.5%.

We perform various experiments on the HepTh dataset [6] to show the effectiveness of all the similarity functions while predicting top 5,10 and 15% of edges of the test data and we show the results obtained in table 2. In the case of HepTh dataset, we can see that when we predict the most important links possible to form in the future, RAE outperforms the rest and there is an increase of 3.5% when we are predicting top 5% edges using RAE when compared to RA. Using CNE we observe that there is an increase of 5% when we are predicting top 15% edges compared to CN.

4 Conclusions and Future Work

From the results we can conclude that our approximations of the state-of-the-art local neighborhood similarity functions perform better than the original similarity functions in terms of classification accuracy. Also common edges between high degree nodes are not so useful in predicting new links. Thus, we completely ignore the contributions of common edges between high degree nodes. We would like to design better similarity schemes that exploit the power law distribution of clique sizes; we would like to consider the contributions of bigger size (more than size 3) cliques. We would also like to make use of more sophisticated clustering and supervised learning approaches to solve this problem and improve the approach of link prediction in terms of classification accuracy.

References

1. Nowell, L., Kleinberg, J.: The Link Prediction Problem for Social Networks. In: CIKM, pp. 556–559. ACM Press, New York (2003)
2. Zhou, T., Lu, L.: Link Prediction in Complex complex networks. J. Phys. A 390, 1150–1170 (2011)
3. Newman, M.E.J.: Networks: An Introduction. Oxford University Press, Oxford (2010)
4. Zhou, T., Lu, L., Zhang, Y.-C.: Predicting Missing links via local information. J. Eur. Phys. B 71, 623–630 (2009)
5. Soundarajan, S., Hopcroft, J.: Using Community information to Improve the Precision of Link Prediction Methods. In: WWW, pp. 607–608. ACM Press, New York (2012)
6. Leskovec, J., Kleinberg, J.: Faloutsos: Graph Evolution: Densification and shrinking Diameters. Technical report, arXiv.org (2007)

A Hybrid Algorithm for the Permutation Flow Shop Scheduling Problem

Arindam Chakravorty[1] and Dipak Laha[2]

[1] Department of Information Technology, St. Thomas' College of Engineering and Technology, Kolkata 700023, India
[2] Department of Mechanical Engineering, Jadavpur University, Kolkata 700032, India

Abstract. This paper considers the application of IA for the classic permutation flow shop scheduling problem. We present a hybrid version of constructively built IA combining with the SA for the n-job, m-machine permutation flow shop scheduling problem to minimize makespan. Based on all the Taillard's benchmark problems, the computational results suggest that the proposed method is very competitive with the existing methods in the literature.

Keywords: Scheduling, Immune algorithm, Simulated annealing, Permutation flow shop, Makespan, Shift neighborhood search.

1 Introduction

In recent times, with the emergence of a variety of evolutionary algorithms, such as GA, ACO, PSO, and DE, another important technique is the immune algorithm (IA), which is guided by the principles concerning the concepts and functions extracted from theoretical immunology. Dasgupta [1], and Hart and Timmis [2] have given an in depth overview of this algorithm. This paper suggests an improved IA by incorporating construction based heuristic, neighborhood solutions and suitable operators of the algorithm. To improve the IA, we hybridize it with SA. Since the flow shop scheduling problem is known to be NP-hard [3], approximate algorithms including computational intelligence based algorithms provide near-optimal/optimal solutions, especially on more difficult, complex search spaces.

Noteworthy research on application of IA to flow shop scheduling has been carried out by Zandieh et al. [4], Kai and Gen-ke, [5], Kurz and Askin [6], Bagheri et al. [7], and Engin and Dyen [8]. Applications of ACO to flow shop scheduling have been studied by Rajendran and Ziegler [9], MMAS [10], Yang and Liao [11], Ahmadizar and Barzinpour [12], and Ahmadizar [13]. In this paper, we propose a hybrid IA combining with SA for the permutation flow shop scheduling problem to minimize the makespan.

2 Problem Definition

We denote p_{ij} as the processing time for each job $i, i\varepsilon n$ on each machine $j, j\varepsilon m$. The objective is to minimize the makespan for permutation flow shop scheduling.

P. Maji et al. (Eds.): PReMI 2013, LNCS 8251, pp. 745–750, 2013.

Let Π denote all permutations of n-job schedules in the solution search space for the flow shop problem. Let π represent a job permutation of schedule, $\pi \varepsilon \Pi$ and $\pi = \{\pi_1, \pi_2, \ldots, \pi_n, \}$. Then, the completion time of π_j on machine m, $C(\pi_j, m)$ is given as: $C(\pi_j, m) = max\{C(\pi_{j-1}, k), C(\pi_j, k-1) + p_{\pi_j,k}\} \ldots (1)$

Then, the makespan of a schedule of n-job is defined as: $MS(\pi) = C(\pi_n, m) \ldots (2)$

Therefore, for $n/m/prmu/C_{max}$ problem, the objective is to determine a permutation schedule $\pi^* \varepsilon \Pi$ such that: $MS(\pi^*) = min_{\pi \varepsilon \Pi} MS(\pi) \ldots (3)$

3 Proposed Algorithm

The proposed algorithm has three major phases, namely, initial solution phase, construction phase and improvement phase. In the initial solution phase, an initial sequence of jobs is created. Construction phase is used to build the sequence of n jobs incrementally by selecting jobs from the initial sequence using IA. Finally, in improvement phase, the solution is further improved by employing neighbourhood search and SA.

Phase 1: Initial sequence generation - The jobs are arranged in descending order of their total processing time (TPT= $\sum_{j=1}^{m} p_{ij}$)to create an initial (seed) sequence.

Phase 2: Construction procedure

Algorithm of Construction phase:

1. Generate an initial sequence of n jobs using phase 1 . Call it seed sequence.
2. Select first r jobs and permute them in r! sequences. Select best M, ($M \leq r!$) sequences on the basis of makespan values from these sequences.
3. Set k = r + 1. Repeat steps 5 to 14 while $k \leq n$.
4. For each of the sequences selected in step 2 repeat step 6.
5. Pick up the k^{th} job from the seed sequence and place it in k positions of the sequence.
6. Sort the k x M sequences of k jobs each obtained in step 6 in ascending order of makespan and select the best M sequences of k jobs each.
7. Clone the M sequences based on Roulette wheel selection principle and create M clone sequences.
8. For each of the M clone sequences use partially mapped crossover with a randomly selected clone from the rest (M - 1) clones. This creates a 2M strong clone pool.
9. For p = 1 to p = 2M repeat steps 11 and 12.
10. Pairwise mutate clone and call the new sequence as newsequence. If makespan of newsequence is less than makespan of clone then set clone = newsequence.
11. Inverse mutate p^{th} clone (clone) and call the new sequence as newsequence. If makespan of newsequence is less than makespan of clone then set clone = newsequence.

12. Select best M clones on the basis of makespan from the pool of 2M clones. Set k = k + 1.

Parameters selection in construction phase: The value of r has been optimized to 4. The algorithm has been tested with M = 10, 12, 15, 18, 20 and 24 and optimized to M = 20.

Phase 3: Improvement procedure

Algorithm of Improvement phase:

1. The best sequence obtained in the construction phase is selected as the best-so-far sequence (best) with makespan=MS_{best}.

2. Let s represent the current sequence, and MS_s the corresponding makespan.

3. Set $s \leftarrow best$ and $MS_s \leftarrow MS_{best}$.

4. Initialize max-count and the start temperature (T_1). Set $count = 1$.

5. Apply the forward shift mechanism to s, separately n-1 times (each case being independent of the others), by randomly picking, in each case, a pair of jobs at two distinct positions h and i (where $1 \leq h < i \leq n$). This produces n-1 neighboring sequences. Find the best of these sequences s', with a makespan of $MS_{s'}$.

6. If $MS_{s'} < MS_s$ then {

6.1 Set the new sequence as the current sequence (that is, set $s \leftarrow s'$ and $MS_s \leftarrow MS_{s'}$);

6.2 If $MS_{s'} < MS_{best}$ then set $best \leftarrow s'$ and $MS_{best} \leftarrow MS_{s'}$; }

7.Apply the backward shift mechanism to s, separately n-1 times (each case being independent of the others), by randomly picking, in each case, a pair of jobs at two distinct positions h and i (where $1 \leq h < i \leq n$). This produces n-1 neighboring sequences. Find the best of these sequences s', with a makespan of $MS_{s'}$.

8.If $MS_{s'} < MS_s$ then {

8.1 Set the new sequence as the current sequence (that is, set $s \leftarrow s'$ and $MS_s \leftarrow MS_{s'}$);

8.2 If $MS_{s'} < MS_{best}$ then set $best \leftarrow s'$ and $MS_{best} \leftarrow MS_{s'}$;}

Else {

8.3 With probability $\exp((MS_s - MS_{s'})/(MS_s * T_1))$ set $s \leftarrow s'$ and $MS_s \leftarrow MS_{s'}$ }

9. Set $count = count + 1$, $T_1 = 0.9 * T_1$. If $count \leq$ max-count go to step 5

Output: the best-so-far sequence (best) and its makespan, MS_{best} as the final solution.

Parameters in improvement phase are taken as: max-count = 200, 250, 300, 500, 800, 1000 and optimized to max-count = 500. T_1 = 100, 150, 200, 250, 300 and optimized to T_1 = 200.

4 Computational Results

The proposed hybrid immune algorithm has been coded in Java and run on Intel Atom 1.6 GHz with 2 GB memory PC. To evaluate the solution quality and

efficiency of the proposed method, the experiments are carried out based on Taillard's benchmark problems [14]. For each problem instance of a problem size, the integer processing time matrix is generated following a uniform discrete distribution [1,99] for each known random number seed value given by Taillard. For each combination of job and machine, 10 independent problem instances are created, producing 120 problem instances in all. To assess the performance of proposed IA with other algorithms, average relative percentage deviation (AR) and the CPU time are used. AR is defined as: $AR = \frac{\sum_{i=1}^{NP}(MS_i - UB_i)/UB_i}{NP} \text{x} 100 \ldots (4)$

where, MS_i and UB_i respectively represent the makespan obtained using an algorithm and the corresponding upper bound for a particular seed value as given in Taillard [14]. The upper bound (UB) is the best value close to the optimal makespan as good as possible and is given by Taillard. Due to the stochastic nature, the proposed algorithm is run for five trials instead of one trial for each problem instance of a problem size and the best and mean values are computed. Then, AR for a problem size is computed using Equation (4), considering each of best and mean makespan values with respect to the UB values.

The proposed algorithm is compared with M-MMAS [9], PACO [9], ACS [11], ACA [12], and NACA [13]. It may be noted that the above mentioned algorithms due to their probabilistic nature have been applied for each problem instance for five trials and the best and mean values have been chosen. Also, the corresponding average CPU time (in seconds) required for these algorithms has been reported. The detailed results are found in Ahmadizar [13]. The above existing algorithms were run on a Pentium 4, 2GHz PC which is slightly faster or at least same as our computer.

The comparative AR_{best}, AR_{mean}, and CPU times of the proposed IA with and without SA and the five competing algorithms are reported in Tables 1 and 2. The computational results reveal that the IA with SA performs significantly

Table 1. Comparison with NACA, M-MMAS and PACO (keeping M-MMAS and PACO same CPU time as that of NACA)

		IA					NACA			M-MMAS		PACO	
		Without SA		With SA									
n	m	AR_{best}	AR_{mean}	AR_{best}	AR_{mean}	$Time(s)$	AR_{best}	AR_{mean}	$Time(s)$	AR_{best}	AR_{mean}	AR_{best}	AR_{mean}
20	5	0.669	1.339	0.172	0.593	0.43	0.000	0.130	0.84	0.127	0.306	0.000	0.138
	10	1.817	2.716	0.486	0.958	0.45	0.079	0.392	1.57	0.447	0.747	0.181	0.458
	20	1.606	2.472	0.536	0.867	0.56	0.102	0.321	3.61	0.315	0.552	0.111	0.280
50	5	0.212	0.448	0.072	0.209	1.47	0.011	0.070	5.03	0.057	0.093	0.053	0.076
	10	1.294	2.364	0.403	1.176	2.19	0.257	0.742	11.14	0.843	1.232	0.761	1.186
	20	2.916	4.054	1.333	1.925	3.51	1.252	1.808	22.71	1.561	2.040	1.781	2.165
100	5	0.127	0.262	0.010	0.081	8.40	-0.006	0.076	19.46	0.036	0.094	0.030	0.073
	10	0.933	1.294	0.404	0.660	12.11	0.282	0.557	43.68	0.463	0.583	0.428	0.585
	20	2.322	2.682	0.714	1.143	16.08	0.761	1.287	93.94	1.183	1.489	1.274	1.565
200	10	0.204	0.464	-0.179	0.012	77.34	0.150	0.341	177.39	0.289	0.469	0.275	0.428
	20	0.706	1.052	-0.043	0.181	105.38	0.306	0.744	389.67	0.571	0.809	0.708	0.957
500	20	-0.376	-0.376	-0.794	-0.794	1440	0.230	0.515	2474.4	0.408	0.526	0.490	0.615
Average:		1.036	1.564	0.260	0.584	138.99	0.285	0.582	270.29	0.525	0.745	0.508	0.711

Table 2. Comparison with ACS and ACA

		IA			ACS		ACA	
		Without SA	With SA					
n	m	AR_{best}	AR_{best}	$Time(s)$	AR_{best}	$Time(s)$	AR_{best}	$Time(s)$
20	5	0.669	0.172	0.43	1.19	3.67	0.368	0.44
	10	1.817	0.486	0.45	1.70	4.00	0.831	0.50
	20	1.606	0.536	0.56	1.60	5.33	0.944	0.63
50	5	0.212	0.072	1.47	0.43	14.67	0.085	2.77
	10	1.294	0.403	2.19	1.89	18.00	1.241	3.73
	20	2.916	1.333	3.51	2.71	24.33	1.990	5.91
100	5	0.127	0.010	8.40	0.22	54.33	0.070	14.15
	10	0.933	0.404	12.11	1.22	65.67	1.059	21.93
	20	2.322	0.714	16.08	2.22	88.00	1.833	37.79
200	10	0.204	-0.179	77.34	0.64	275.33	0.434	141.52
	20	0.706	-0.043	105.38	1.30	631.67	1.236	254.06
500	20	-0.376	-0.794	1440	1.68	5133.00	1.444	3744.3
Average:		1.036	0.260	138.99	1.40	526.50	0.961	352.31

better than all the algorithms except NACA where it shows comparable solution. However, the IA without SA produces inferior solutions compared to other methods.

5 Conclusion

In this paper, we present a hybrid version of the immune algorithm and the simulated annealing to enhance the quality of solution while retaining the same order of computational time. The proposed algorithm employs hybridization of two evolutionary computational methods which has enabled better searching options. This approach has shown to yield results that are better or competitive than the other notable evolutionary algorithms.

References

[1] Dasgupta, D.: Artificial immune systems and their applications. Springer (1999)
[2] Hart, E., Timmis, J.: Applications areas of AIS: The past, the present and the future. Applied Soft Computing 8, 191–201 (2008)
[3] Gonzalez, T., Sahani, S.: Flow shop and job shop scheduling: complexity and approximation. Operations Research 26, 36–52 (1978)
[4] Zandieh, M., Fatemi Fhomi, S.M.T., Moattar Husseini, S.M.: An immune algorithm approach to hybrid flow shops scheduling with sequence-dependent setup times. Applied Mathematics and Computation 180, 111–127 (2006)
[5] Kai, S., Gen-ke, Y.: Hybrid artificial immune system and extremal optimization algorithm for permutation flowshop scheduling problem. Journal of Shanghai Univ (Engl Ed) 12(4), 352–357 (2008)

[6] Kurz, M.E., Askin, R.G.: Scheduling flexible flow lines with sequence-dependent setup times. European Journal of Operational Research 159, 66–82 (2004)
[7] Bagheri, A., Zandieh, M., Mahdavi, I., Yazdani, M.: An artificial immune algorithm for the flexible job-shop scheduling problem. Future Generation Computer Systems 26, 533–541 (2010)
[8] Engin, O., Doyen, A.: A new approach to solve hybrid flow shop scheduling problems by artificial immune system. Future Generation Computer Systems 20, 1083–1095 (2004)
[9] Rajendran, C., Ziegler, H.: Ant-colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. European Journal of Operational Research 155, 426–438 (2004)
[10] Stuetzle, T.: An ant approach for the flow shop problem. Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing (EUFIT) 3, 1560–1564 (1998)
[11] Ying, K.C., Liao, C.J.: An ant colony system for permutation flow-shop sequencing. Computers & Operations Research 31, 791–801 (2004)
[12] Ahmadizar, F., Barzinpour, F.: A hybrid algorithm to minimize makespan for the permutation flow shop scheduling problem. International Journal of Computational Intelligence Systems 3, 853–861 (2010)
[13] Ahmadizar, F.: A new ant colony algorithm for makespan minimization in permutation flow shops. Computers & Industrial Engineering 63, 355–361 (2012)
[14] Taillard, E.: Benchmarks for basic scheduling problems. European Journal of Operational Research 64, 278–285 (1993)

Author Index