

L. Ashok Kumar
L. S. Jayashree
R. Manimegalai *Editors*

Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications

AISGSC 2019

 Springer

Proceedings of International Conference on Artificial
Intelligence, Smart Grid and Smart City
Applications

L. Ashok Kumar • L. S. Jayashree
R. Manimegalai
Editors

Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications

AISGSC 2019

 Springer

Editors

L. Ashok Kumar
Department of Electrical and Electronics
Engineering
PSG College of Technology
Coimbatore, Tamil Nadu, India

L. S. Jayashree
Department of Computer Science and
Engineering
PSG College of Technology
Coimbatore, Tamil Nadu, India

R. Manimegalai
Department of Information Technology
PSG College of Technology
Coimbatore, Tamil Nadu, India

ISBN 978-3-030-24050-9

ISBN 978-3-030-24051-6 (eBook)

<https://doi.org/10.1007/978-3-030-24051-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The genesis of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications (**AISGSC-2019**) was started during October 2017, with the aim of bringing people working in research areas related to the circuit branches and computing branches together and at the same time helping researchers publish their research work in quality forums such as Scopus indexed journals. Departments of Electrical and Electronics Engineering, Computer Science and Engineering and Information Technology, PSG College of Technology decided to jointly organize the International Conference on Artificial Intelligence, Smart Grid and Smart City Applications (AISGSC 2019) during 3–5, January 2019.

We, the Organizing Secretaries, believed that we could rope in good-quality journals to publish the work done by researchers and we were able to do so during the process of organizing this conference. We feel happy to provide a forum to communicate and discuss various research findings and innovations in Artificial Intelligence, Smart Grid, Smart City and related areas. We believe that AISGSC 2019 provides great scope for researchers to identify new research ideas from papers presented during various technical sessions and pursue research issues and challenges in the above-mentioned areas.

Smart Cities Mission is an urban renewal and retrofitting program by the Government of India with the aim to develop 100 cities across the country making them citizen friendly and sustainable. The Union Ministry of Urban Development is implementing the mission in collaboration with the State Governments of the respective cities. *Smart Cities Mission* aims to develop areas nearby and within 100 cities in the country as model areas using an area development plan, which is expected to have a rub-off effect on other parts of the city, and nearby cities and towns. The range of smart city applications is highly diverse with several verticals such as smart power, smart transport, smart buildings, smart health-care and smart security. The major functions to focus on in implementing smart-city applications are: *i*) creation of smart platform, *ii*) monitoring and basic analytics, *iii*) deep analytics, *iv*) smart control, *v*) instant interaction with user via applications and

vi) integrating several solutions into a single bundle. Therefore, in AISGSC 2019, it was decided to give the following themes in call for papers:

- Electrical Power Systems Engineering
- Communication Systems Engineering
- Control Systems and Wearable electronics
- Artificial Intelligence and Cognitive Computing
- IoT and Cyber Physical Systems
- Text, Image and Video Analytics
- Networking and Cyber Security
- High-Performance and Distributed Systems
- Embedded & VLSI Systems and Applications

AISGSC 2019 is conducted in collaboration with the Asian Institute of Technology, Thailand, and the National Institute of Technology, Indonesia, and technically supported by Springer, the Institution of Electronics and Telecommunication Engineers, Indian Association of Energy Management Professionals, Indian Green Building Council, Computer Society of India (Coimbatore Chapter), the Information Systems Audit and Control Association (ISACA-Coimbatore Chapter) and the Institution of Engineers India (IEI).

We sincerely appreciate the cooperation and guidance received from members of the organizing committee, national and international advisory committees and technical committee. We place on record our sincere thanks to the corresponding Heads of Departments, who are the conveners of this conference, for their time, effort, and enthusiastic guidance in planning and organizing this conference. We wish to express our gratefulness and thanks to our Principal **Dr. R Rudramoorthy**, who was the driving force behind this conference. Our heartfelt thanks to our Managing Trustee **Shri. L Goplakrishnan** for his continuous support and suggestions for improvement.

We are greatly indebted to our co-organizing secretaries and coordinators for their tireless support right from the planning stage until the curtain-closure stage of this mega event. We thank all the organizing committee members, faculty, staff and students belonging to departments of EEE, CSE and IT for their wholehearted involvement and cooperation extended in organizing this conference. Our special thanks are due to the reviewers for sparing their valuable time in the arduous task of reviewing the papers and suggesting improvements in submitted papers. We express our sincere thanks to all guests, speakers and chairpersons for various sessions during 3–5 January 2019. We thank all the in-land and foreign authors for showing a great interest in sending their papers to this conference. We thank all the registered participants for their immense confidence and belief in the institution's great reputation and brand name.

Our special thanks to the Placement Cell of our institution for their magnificent support in mobilizing sponsorship from our leading recruiters. We also acknowledge with gratitude, the financial support provided to the conference by Cognizant Technology Solutions, TNSCST, IE(I), Advanced Digital Micro Logic and DST ASEAN. We also thank all sponsors who supported us in exhibiting their products

and services. We profusely thank the print, electronic media and Okulus Digital for the wide coverage and publicity of our conference.

We received an overwhelming response of more than 600 papers submitted for the conference. A total of 250 papers were selected for oral and poster presentations after a thorough double blind review process. Among the selected 250 papers, 90 papers were recommended for publication by Springer, USA, and 75 papers were recommended for publication in various Scopus indexed journals. As part of this conference, five workshops in the following topics have been planned for the benefit of research, student and teaching communities:

- Machine Learning and Data Analytics for Smart Grid Applications
- AI, Machine Learning and Big Data Analytics for Smart City Applications
- IoT and Edge/Fog Analytics for Smart City Applications
- Smart Grid Testbed from Typhoon HIL
- Editorial Workshop: Publishing Scholarly Books and Journal Papers

Apart from inaugural and valedictory sessions, 10 plenary talks by renowned resource persons from reputed academic institutions and industries and 28 paper presentation sessions have been scheduled. A day-long product demo sessions from various industries are also scheduled.

We are sure that this conference, a result of the meticulous efforts by various organizing committees, will help the participants to explore new avenues and dimensions pertaining to the central theme of the conference, namely Artificial Intelligence, Smart Grid and Smart City Applications. The recommendations given by various experts involved in the conference programme will be summarized and communicated to various departments of the government, public and private sector organizations for their best possible use in implementing Smart Cities in India.

Coimbatore, Tamil Nadu, India

L. Ashok Kumar
L. S. Jayashree
R. Manimegalai
Organizing Secretaries (AISGSC-2019)

Contents

1	Fractional-Order PID Controller Optimized by SCA for Solar System	1
	Raj Kumar Sahu, Binod Shaw, and Jyoti Ranjan Nayak	
2	LVRT Capability Improvement in a Grid-Connected DFIG Wind Turbine System Using Neural Network-Based Dynamic Voltage Restorer	11
	Arun Kumar Puliyadi Kubendran and L. Ashok Kumar	
3	Detection and Classification of Power Quality Events Using Wavelet Energy Difference and Support Vector Machine	21
	Arun Kumar Puliyadi Kubendran and L. Ashok Kumar	
4	PMSM Drive Using Predictive Current Control Technique for HVAC Applications	31
	L. Ashok Kumar and V. Indragandhi	
5	Grid-Connected 5 kW Mono-crystalline Solar PV System	43
	L. Ashok Kumar, Sheeba Babu, and V. Indragandhi	
6	An Add-in Tool for BIM-Based Electrical Load Forecast for Multi-building Microgrid Design	57
	Jasim Farooq, Rupendra Kumar Pachauri, R. Sreerama Kumar, and Paawan Sharma	
7	An Investigation on Torque Ripple Minimization of Switched Reluctance Motor Using Different Power Converter Topologies Using Intelligent Techniques	69
	M. Gengaraj, L. Kalaivani, K. Cherma Jeya, P. Eswari Prabha, A. M. Kirthika, and M. Vavuniya	

8	Design of Half-Ring MIMO Antenna to Reduce the Mutual Coupling	81
	K. Vasu Babu and B. Anuradha	
9	Optimal Allocation of Distributed Generation Using Clustered Firefly Algorithm	89
	K. Banumalar, B. V. Manikandan, and S. Sundara Mahalingam	
10	CDM-Based Two-Degree-of-Freedom PID Controller Tuning Rules for Unstable FOPTD Processes	99
	Somasundaram S and Benjanarasuth T	
11	Real-Time Energy Management System for Solar-Wind-Battery fed Base Transceiver Station	109
	W. Margaret Amutha and V. Rajini	
12	IOT-Based Adaptive Protection of Microgrid	123
	O. V. Gnana Swathika and K. T. M. U. Hemapala	
13	Performance Comparison Between Sensor and Sensorless Control of Permanent Magnet Synchronous Motor with Wide Speed Range of Operations	131
	N. Krishna Kumari, D. Ravi Kumar, and K. Renu	
14	Passive Fault-Tolerant Control Based on Interval Type-2 Fuzzy Controller for Coupled Tank System	145
	Himanshukumar R. Patel and Vipul A. Shah	
15	Enhanced Isolated Boost DC–DC Converter with Reduced Number of Switches	155
	Anjel J and Gerald Christopher Raj I	
16	Harmonic Intensity Reduction Technique for Three Phase VSI Drive through Double Randomness	163
	P. Arulkumar, K. Jaiganesh, and N. P. Subramaniam	
17	PV-Based Multilevel Inverter-Fed Three-Phase Induction Motor with Improved Time and Speed of Response	173
	Chandrasekaran S and Durairaj S	
18	Adaptive Disturbance Observers for Building Thermal Model	185
	Mallikarjun Soudari, Seshadhri Srinivasan, and B. Subathra	
19	HTSA Optimized PID-Based MPPT for Solar PV System	197
	Shashikant and Binod Shaw	
20	Performance Analysis of UFMC System with Different Prototype Filters for 5G Communication	207
	M. Maheswari, N. R. Nagarajan, and M. Banupriya	

21 Fully Convolved Neural Network-Based Retinal Vessel Segmentation with Entropy Loss Function 217
 V. Sathananthavathi, G. Indumathi, and A. Swetha Ranjani

22 Solar Power Forecasting Using Adaptive Curve-Fitting Algorithm 227
 N. Sampathraja, L. Ashok Kumar, R. Saravana Kumar, and I. Made Wartana

23 A Review of Electric Vehicle Technologies 237
 P. Ravi Kumar, C. Gowri Shankar, R. Uthirasamy, and V. J. Vijayalakshmi

24 Gabor Filter-Based Tonsillitis Analysis Using VHDL 247
 P. Nagabushanam, S. Thomas George, D. S. Shylu, and S. Radha

25 Incorporation of Modified Second-Order Adaptive Filter in MFGCI for Harmonic Mitigation of Microgrid 259
 P. C. Keerthiga, G. Gabriel Santhosh Kumar, and S. Hemila Haland

26 Optimal DAU Placement for Smart Distribution Grid Communication Network 269
 S. Premkumar, M. Susithra, and V. Saminadan

27 Long-Term Forecasting of Hybrid Renewable Energy Potential Using Weibull Distribution Method in Coimbatore 279
 Anuradha J, Soundarrajan A, and Rajan Singaravel M M

28 Efficient and Improved ANN-Based Voltage-Rise Mitigation Strategy in Distribution Network with Distributed Solar Photovoltaic System 289
 Neenu Thomas, R. Jayabarathi, and T. N. P. Nambiar

29 Control of Buck Converter by Fuzzy Controller for Wind Energy: Battery System 299
 Sheeba Babu and L. Ashok Kumar

30 A Survey on Secure Beamforming in MIMO-NOMA-Based Cognitive Radio Network 311
 Thulasimani Lakshmanan and Hyils Sharon Magdalene Antony

31 Hybrid Optimization of Cuckoo Search and Differential Evolution Algorithm for Privacy-Preserving Data Mining 323
 J. Sudeeptha and C. Nalini

32 Using Sliding Window Algorithm for Rainfall Forecasting 333
 M. Vijaya Chitra and Grasha Jacob

33	Air Pollution-Level Estimation in Smart Cities Using Machine Learning Algorithms	343
	M. Nelgadevi and Grasha Jacob	
34	Implicit Continuous User Authentication Using Swipe Actions on Mobile Touch Screen with ANN Classifier	353
	Christy James Jose and M. S. Rajasree	
35	A Review on Graph Analytics-Based Approaches in Protein-Protein Interaction Network	365
	D. Narmadha, A. Pravin, G. Naveen Sundar, and Premnath Dhanaraj	
36	A Survey on Emotion Detection Using EEG Signals	375
	Oshin R. Jacob and G. Naveen Sundar	
37	A Smart Agricultural Model Using IoT, Mobile, and Cloud-Based Predictive Data Analytics	383
	P. Anand Prabu and L. S. Jayashree	
38	Machine Translation Using Deep Learning: A Comparison	389
	S. Swathi and L. S. Jayashree	
39	Societal Impact of Framework for Energy-Efficient Clustering Algorithms in Mobile Wireless Sensor Networks	397
	K. Juliet Catherine Angel and E. George Dharma Prakash Raj	
40	Energy Demand Prediction Using Linear Regression	407
	T. Manojpraphakar and Soundarrajan A	
41	Risk Prediction Analysis of Cardiovascular Disease Using Supervised Machine Learning Techniques	419
	A. Ishwarya and S. K. Jayanthi	
42	Safest Secure and Consistent Data Services in the Storage of Cloud Computing	433
	Geethu Mary George and L. S. Jayashree	
43	Agile Supply Chain Management Enabled by the Internet of Things and Microservices	449
	G. Selvakumar and L. S. Jayashree	
44	Production and Characterization of Bio-Fertilizers from Tree Leaves Utilizing an Automated Hot Composting Chamber with Cyber-Physical Systems	457
	Mahendran Rajagopalan, Vijayakumar Arumugam, Uma Dharmaligam, KaviniLavuvu Anbalagan, and Anupriya Chandrasekaran	

45 Spectrum Sensing Based on Cascaded Approach for Cognitive Radios 467
 N. Iswarya and L. S. Jayashree

46 Remote Process Monitoring and Control Through IIoT 481
 Swetha R Kumar, Sangavarthini C.S., and L Ashok Kumar

47 Case-Based Reasoning (CBR)-Based Smart Indoor Navigation 491
 G. R. Karpagam, K. Eshwar, K. Karthikeyan, and M. Syed Hameed

48 A Survey on Medical Image Registration Using Deep Learning Techniques 505
 M. C. Shunmuga Priya and L. S. Jayashree

49 Agent-Based Temperature Monitoring System 513
 S. Jaswanth and L. S. Jayashree

50 Classification of Phonemes Using EEG 521
 R. Aiswarya Priyanka and G. Sudha Sadasivam

51 Attribute Table-Based Multipath Routing Protocol to Improve Network Lifetime in Multi-hop WSN 531
 B. Sherin, M. Senthil Vadivu, and A. Ayub Khan

52 Application of Subjective and Objective Integrated Weightage (SOIW) Method for Decision-Making (MADM) in Distribution System 541
 Sachin Gorakh Kamble, Kinhal Vadirajacharya, and Udaykumar Vasudeo Patil

53 Visual Importance Identification of Natural Images Using Location-Based Feature Selection Saliency Map 557
 Malayil Abhayadev and T Santha

54 Missing Values and Class Prediction Based on Mutual Information and Supervised Similarity 573
 Nagalakshmi K. and S. Suriya

55 Fake Product Review Detection and Removal Using Opinion Mining Through Machine Learning 587
 Minu Susan Jacob, Selvi Rajendran, V. Michael Mario, Kavali Tejasri Sai, and D. Logesh

56 Ask Less: Scale Market Research Without Annoying Your Customers 603
 Venkatesh Umaashankar and S. Girish Shanmugam

57	Preferential Resource Selection and Scheduling of Cloud Resources Pivot on Value of Information	613
	Renu Suresh Ganvir, Salaja Silas, and Elijah Blessing Rajsingh	
58	A Survey on Supervised and Unsupervised Learning Techniques	627
	K. Sindhu Meena and S. Suriya	
59	Performance Study of IPv6/IPv4 MANET (64MANET) Architecture	645
	S. Manimozhi and J. Gnana Jayanthi	
60	Internet of Things: A Technical Perspective Survey	659
	S. Margaret Amala and J. Gnana Jayanthi	
61	Analysis on DGHV and NTRU Fully Homomorphic Encryption Schemes	669
	B. Santhiya and K. Anitha Kumari	
62	Automated Image Captioning for Flickr8K Dataset	679
	K. Anitha Kumari, C. Mouneeshwari, R. B. Udhaya, and R. Jasmitha	
63	RAkEL Algorithm and Mahalanobis Distance-Based Intrusion Detection System Against Network Intrusions	689
	R. Padmashani, M. Nivaashini, and R. Vidhyapriya	
64	Vaguely Node Classification Scheme for Wireless Networks to Design an Intrusion Detection System	697
	S. Latha and V. Sinthu Janita Prakash	
65	Dynamic Traffic Light Scheduling for Emergency Vehicles Using Fog Computing	709
	S. Sarathambekai, T. Vairam, and A. Dharani	
66	Cloud Database – A Technical Review	721
	S. Sakthivel and J. Gnana Jayanthi	
67	Projection of Population and Prediction of Food Demand Through Mining and Forecasting Techniques	733
	J. Antonita Shilpa and V. Bhanumathi	
68	Detection of Hairline Fracture Foot Using Canny Operator and Wavelet Packet Transform	741
	D. S. Karthika, K. S. Biju, G. H. Silpa, and C. Girish Kumar	
69	Image Encryption-Then-Compression System for Secure Transmission via Hybrid Henon Chaotic Map	749
	P. Sridevi and J. Suguna	

70 Analysis of Primary Emulsion Attack in Cognitive Radio Using Distributed On-Demand Routing Protocol 757
 Neelaveni Rangaraj and Sridevi Balu

71 Heart Disease Prediction Using Retinal Fundus Image 765
 R. Rekha, V. P. Brintha, and P. Anushree

72 Blind Speech Enhancement Using Adaptive Algorithms 773
 P. Shanmuga Priya and S. Selva Nidhyananthan

73 Android Malware Detection 781
 Shymala Gowri Selvaganapathy, G. Sudha Sadasivam, Hema Priya N, Rajeshwari N, Dharani M, and K. Karthik

74 Test Data Compression Methods: A Review 791
 S. Rooban and R. Manimegalai

75 Piripori: Morphological Analyser for Tamil 801
 M. Suriyah, Aarthi Anandan, Anitha Narasimhan, and Madhan Karky

76 A Comprehensive Survey on Strategies in Multicore Architectures, Design Considerations and Challenges 811
 R. Radhika, N. Anusha, and R. Manimegalai

77 Blockchain-based e-Voting as a Service 819
 R. S. Shyam Prakash and G. R. Karpagam

78 Optimum Resource Allocation Techniques for Enhancing Quality of Service Parameters in Cloud Environment 831
 M. Kandan and R. Manimegalai

79 MIPGIOT: Monitoring and Improving the Productivity in Garment Unit Using IOT 841
 V. G. Prabhu and R. Manimegalai

80 Text and Audio Transfer Using LI-FI Technology 847
 S. Sabareeswaran, G. Madumitha, and M. Shruthi

81 Hyperspectral Image Segmentation Using Evolutionary Multifactorial Spectral Analysis for OMEGA Dataset 861
 Nagarajan Munusamy and Rashmi P. Karchi

82 Novel Lifting Filter Bank for Bird Call Analysis 871
 N. Subbulakshmi and R. Manimegalai

83 Automatic Classification of Solid Waste Using Deep Learning 881
 V. P. Brintha, R. Rekha, J. Nandhini, N. Sreekaarthick, B. Ishwaryaa, and R. Rahul

84 Relevancy and Similarity Aware Drug Comment Classification Framework on Social Media Drug Posts 891
D. Krithika Renuka and B. Rosiline Jeetha

85 Enhanced Particle Swarm Optimization with Genetic Algorithm and Modified Artificial Neural Network for Efficient Feature Selection in Big Data Stream Mining 909
S. Meera and B. Rosiline Jeetha

86 Feature Selection Techniques for Email Spam Classification: A Survey 925
V. Sri Vinitha and D. Karthika Renuka

87 A Novel Paradigm Towards Exploration of Rechargeable WSN Through Deep Learning Architecture for Prolonging Network Lifetime 937
M. Ezhilarasi and V. Krishnaveni

88 Why Feature Selection in Data Mining Is Prominent? A Survey 949
M. Durairaj and T. S. Poornappriya

Abbreviations

A	Actual Value
AA	Average Accuracy
AABC-ANN	Acceleration Artificial Bee Colony–Artificial Neural Network
ABC	Artificial Bee Colony
AC	Arithmetic Coder
ACF	Adaptive Curve Fitting
ACLMessage	Access Control List Message
ACM	Association of Computing Machinery
ACO	Ant Colony Optimization
ADE	Adaptive Disturbance Estimator
ADE	Adverse Drug Events
AdOp	Advertisement Output
AENS	Average Energy Not Supplied
AES	Advanced Encryption Standard
AEV	All Electric Vehicles
AF	Amplify and Forward
AHP	Analytic Hierarchy Process
AI	Asymmetry Index
AIC	Akaike Information Criterion
AIS-DAG	Artificial Immune System-Directed Acyclic Graph
ALU	Arithmetic Logic Unit
AMQP	Advanced Message Queuing Protocol
ANF	Adaptive Notch Filter
ANM -	Asymmetric Numerical Method
ANN	Artificial Neural Network
AODV	Ad Hoc On Demand Distance Vector
AOL	America Online
APC	Adaptive Protection Center

API	Application Programming Interfaces
APK	Android Application Package (which is a dataset)
AQI	Air Quality Index
AR	Average Rainfall
ARFF	Attribute-Relation File Format
ASD	Adjustable Speed Drive
AT&T	American Telephone & Telegraph Company
ATE	Automatic Testing Equipment
ATEQMR	Attribute Table Based Energy Efficient QoS Multipath Routing
ATPG	Automatic Test Pattern Generator
AUC	Area Under Curve
AutoDR	Automated Demand Response
AWGN	Additive White Gaussian Noise
B2C	Business to Customer
BaTS	Budget Constraint Scheduling
BCTVB	Building Controls Virtual Test Bed
BDA	Building Design Advisor
BDMS	Bi-directional Mapping System
BER	Bit Error Rate
BEV	Battery Electric Vehicle
BGP	Border Gateway Protocol
BIA	Bump In the API
BIC	Bayesian Information Criterion
Bio GRID	The Biological General Repository for Interaction Datasets
BIS	Bump In the Stack
BLER	Block Error Rate
BLEU	Bilingual Evaluation Understudy
BMG	Block Memory Generator
BN	Bayesian Network
BoS	Balance of Systems
BRAM	Block Random Access Memory
BROM	Block Read Only Memory
BS	Base Station
BSS	Blind Source Separation
BTS	Base Transceiver Station
BWO	Binary Wolf Optimization
ByER -	Byte-Per-Rate
CAIDI	Customer Average Interruption Duration Index
CAIFI	Customer Average Interruption Frequency Index
CART	Classification and Regression Tree
CBR	Constant Bit Rate

CBR	Case-Based Reasoning
CC	Control Center
CCR	Communication Computational Ratio
CCSPWM	Chaotic Carrier Sinusoidal PWM
CCTree	Categorical Clustering Tree (an algorithm used for automated grouping and classifying malware)
CD	Covariance-Based Detection
CDF	Cohen-Daubach-Feauveau
CDM	Coefficient Diagram Method
CEA	Central Electricity Authority
CFA	Cuttlefish Algorithm
CFFA	Clustered Firefly Algorithm
CFFB	Current-Fed Full Bridge
CFS	Correlation-Based Feature Selection
CH	Cluster Head
CHOL	Cholesterol Level
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CLI	Command Line Interface
CM	Cluster Member
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolution Neural Networks
CoAP	Constrained Application Protocol
CoBF	Coordinated Beamforming
COF	Convection Factor
COPD	Chronic Obstructive Pulmonary Diseases
CP	Chest Pain
CPAD	Cyclic Prefix Autocorrelation Detection
CPCB	Central Pollution Control Board
CPD	Custom Power Device
CPU	Central Processing Unit
CRIST900	Content Retargeting Image Resizing Technique 900
CRN	Cognitive Radio Network
CS	Cuckoo Search
CS	Cloud Server
CS	Chi-Square
CSC	Continuous Seam Carving
CSD	Cyclo Stationary Feature Detection
CSI	Channel State Information
CSP	Cloud Service Providers
CSR	Cyclic Shift Register
CSTR	Continuous Stirred Tank Reactor
CT	Current Transformer
CT	Computed Tomography
CTE	Compression-Then-Encryption

CTLC	Coupled Tank Level Control System
CVD	Cardiovascular Disease
D NOMA	Dynamic Non-Orthogonal Multiple-Access
DaaS	Development as a Service
DAG	Directed Acyclic Graph
DAU	Data Aggregation Unit
Db	Decibels
DaaS	Database as a Service
DBMS	Database Management System
DC	Direct Current
DCDB	DC Distribution Box
DCM	Discontinuous Conduction Mode
DDDA	Dynamic Drug Data Examination
DDoS	Distributed Denial of Service (one of the attack types)
DE	Differential Evolution
DEA	Differential Evolution Algorithm
DEAP	Database for Emotion Analysis Using Physiological Signals
DER	Distributive Energy Resources
DF	Decode and Forward
DFIG	Doubly Fed Induction Generator
DFNLMS	Dual Fast Normalized Least Mean Square
DFT	Design For Testability
DFU	Diabetes Foot Ulcer
DG	Distributed Generators
DGHV	Diji Gentry Halevi Vaikuntanathan
DHT11	Temperature and Humidity Sensor
DIP	Database of Interacting Proteins
DM	Data Mining
DM	Decision Maker
DMAIC	Define, Measure, Analyze, Improve, Control
DMGEECA	Density-Based Mean Grid Energy-Efficient Clustering Algorithm
DNS	Domain Name System
DOD	Depth Of Discharge
DODAG	Destination-Oriented Directed Acyclic Graph
DoS	Denial of Service
DPB	Distribution Panel Board
DRIVE	Digital Retinal Images for Vessel Extraction
DS	Distribution System
DSMIPv6	Dual Stack Mipv6
DSP	Digital Signal Processing
DSTM	Dual Stack Transition Mechanism
DT	Decision Tree
DTC	Direct Torque Control

DVFS	Dynamic Voltage and Frequency Scaling
DVR	Dynamic Voltage Restorer
EC	Electrical Conductivity
EC2	Amazon Elastic Compute Cloud
ECC	Envelope Correlation Coefficient
ECC	Elliptic Curve Cryptography
ECG	Electrocardiogram
ED	Energy Differences
ED	Energy Detection
EDC-AIRS	Evolutionary Data-Conscious Artificial Immune Recognition System
EDLC	Electric Double-Layer Capacitors
EEECA	Enhanced Energy-Efficient Clustering Algorithm
EEG	Electroencephalogram
EFDR	Extended Frequency Directed Run Length
EKF	Extended Kalman Filter
ELECTRE	Elimination and Choice Translating Reality (Elimination And Choice Expressing Reality)
ELM	Extreme Learning Machine (an algorithm used for classification)
EME	Energy with Minimum Eigenvalue Ratio
EMF	Electromotive Force
EMI	Electro Magnetic Interference
EMS	Energy Management System
ENVI	Environment for Visualizing Images
EPA	Environmental Protection Agency
EPF	Energy Pattern Factor Method
EPS	Electric Propulsion System
EPSOGA-MANN	Enhanced Particle Swarm Optimization With Genetic Algorithm-Modified Artificial Neural Network
EQMH	Energy-Efficient and QoS-Based Multipath Hierarchical Routing Protocol
ERA	Efficient Resource Allocation
ERNET	Education and Research Network
ERNN	Elman Recurrent Neural Network
ertPS	Extended Real-Time Polling Service
ESA	Evolutionary Multifactorial Spectral Analysis
ESS	Energy Storage System
ETA	Estimated Time To Arrive
ETC	Encryption-Then -Compression
ETE	End To End
EV	Electric Vehicle
EV	Earlier Variation
EY	Earlier Year

F	F-Score
FAERS	FDA Adverse Event Reporting System
FBMC	Filter Bank Multi-carrier
FBS	Fasting Blood Sugar
FC	Fuel Cells
FCBF	Fast Correlation-Based Feature Selection
FCE	Final Control Element
FCEV	Fuel Cell Electric Vehicle
FCFS	First Come First Serve
FCM	Fuzzy C-Means Clustering
FCN	Fully Convolutional Networks
FDA	Food and Drug Administration
FDR	Frequency-Directed Run Length
FEECA	Framework for Energy-Efficient Clustering Algorithms
FFT	Fast Fourier Transform
FHE	Fully Homomorphic Encryption
FLC	Fuzzy Logic Controller
FN	False Negative
FNN	Feed Forward Neural Network
FOC	Field-Oriented Control
FOSS	Free and Open-Source Software
FOPID	Fractional Order Proportional Integral Derivatives
FOPID	Fractional Order PID
FP	False Positive
FPGA	Field Programmable Gate Array
FPID	Fuzzy Proportional Integrator And Derivative
FPOA	Field Programmable Object Array
FPR	False Positive Rate
FPR	False Positive Rate
FPU	Floating Point Unit
FR	Forecasted Rainfall
FSM	Finite State Machine
FTC	Fault Tolerant Control
GA	Genetic Algorithm
GC	Green Cloud
GCI	Grid-Connected Inverter
GD	Group Delay
GEECA	Grid-Based Energy-Efficient Clustering Algorithm
GF	Gaussian Function
g_i	Stability Indices
GI	Gini Index
GIS	Geographic Information System
GP	Galois Field
GPRS	General Packet Radio Service

GPS	Global Positioning System
GR	Gain Ratio
GRMSE	Geometric Root Mean Square Error
GSC	Grid Side Converter
GSM	Global System for Mobile Communications
GSR	Galvanic Skin Response
GUI	Graphical User Interface
HAN	Home Area Network
HC	Huffman Coder
HDFS	Hadoop Distributed File System
HDL	Hardware Description Language
HDL	High-Density Lipoprotein
HE	Homomorphic Encryption
HEV	Hybrid Electric Vehicle
HHT	Hilbert Huang Transform
HIRDRPWM	Harmonic Intensity Reduction Double Randomness PWM
HPWM	Hybrid Pulse With Modulation
HRES	Hybrid Renewable Energy Sources
HSF	Harmonic Spread Factor
HTSA	Heat Transfer Search Algorithm
HTSVM	Hybrid Transductive Support Vector Machine
I/O	Input/Output
I/P	Current to Pressure
IaaS	Infrastructure as a Service
IC	Integrated Circuits
ICE	Internal Combustion Engines
ICEV	Internal Combustion Engine Vehicle
ICSA	Improved Clonal Selection Algorithms
ICT	Integrated Information and Communication Technologies
ICT	Information and Communication Technology
ID	Intrusion Detection
IDA	Indoor Climate and Energy
IDEA	Improved Differential Evolution Algorithm
IDS	Intrusion Detection System
IETF	Internet Engineering Task Force
IG	Information Gain
IIoT	Industrial Internet Of Things
IMAP	Integrated Memory Array Processor
IOPID	Integer Order Proportional Integral Derivatives
IoT	Internet Of Things
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
IRC	Internet Relay Chat
ISO	International Organization for Standardization

ISP	Internet Service Provider
IT	Information Technology
IT2FLC	Interval Type-2 Fuzzy Logic Control Approach
JADE	Java Agent Development Framework
JB	Junction Box
JD	Joint Director
K	Process Gain
KB	Kilo Byte
K_c	Proportional Gain
K_D	Derivative Gain
K_I	Integral Gain
KMA	Karnik Mendel Algorithm
KNN	K Nearest Neighbour
K_P	Proportional Gain
L	Process Time Delay
LBFSM	Location-Based Feature Selection Saliency Map
LDL	Low-Density Lipoprotein
LED	Light-Emitting Diode
LFSR	Linear Feedback Shift Register
LIC	Level Indicator and Controller
LI-FI	Light Fidelity
LMBP	Levenberg-Marquardt Back Propagation
LMPC	Linear Model Predictive Control
LMS	Least Mean Square
LMT	Logistic Model Tree
LoRaWAN	Long Range Wide Area Network
LQR	Linear Quadratic Regulator
LR	Linear Regression
LReLU	Leaky ReLU
LSTM	Least Short-Term Model
LSTM	Long Short-Term Memory
LT	Level Transmitter
LVRT	Low Voltage Ride Through
LWE	Lattice-Based Encryption
MAC	Machine-Authenticated Code
MAC	Media Access Control
MAC	Multiply Accumulate
MADM	Multiple Attribute Decision Making
MAE	Mean Absolute Error
MAS	Multi-Agent System
MATLAB	Matrix Laboratory
MBCx	Monitoring-Based Commissioning
MC	Moisture Content
MCC	Matthews Correlation Coefficient

MCI	Microwave Communication Commission
MD	Matched Filter Detection
MD5	Message Digest 5
MEV	Mean Earlier Variation
MF	Membership Function
MFGCI	Multifunctional Grid Connected Inverter
MG	Micro Grid
MH-RF	Multi-Hop Radio Frequency
MI	Mutual Information
MIMO	Multiple Input, Multiple Output
MIP-ALG	Mobile Internet Protocol–Application Level Gateway
MIPS	The Munich Information Center for Protein Sequences
MIRT	Mobile Infrared Transmitter
MISO	Multiple Input, Single Output
ML Algo	Machine Learning Algorithm
MLBMPSTO	Multi-Objective Load Balancing Mutation Particle Swarm Optimization
MLE	Maximum Likelihood Estimate
MLI	Multilevel Inverter
MLP	Multi-Layer Perceptron
MM	Majorization Minimization
MM CVX	Majorization Minimization Convex Optimization
MM SCF	Majorization Minimization Semi-Closed Form Secrecy Rate Optimization
MME	Maximum to Minimum Eigenvalue Ratio
MN	Mobile Node
MOABC	Multi-Object Artificial Bee Colony
MOS	Mean Opinion Score
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MOV _s	Metal Oxide Varistors
MPC	Model Predictive Control
MPC	Microgrid Protection Center
MPPT	Maximum Power Point Tracking
MPV	Mean Present Variation
MQTT	Message Queuing Telemetry Transport
MRA	Multi-Resolution Analysis
MRAS	Model Reference Adaptive System
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
MTPA	Maximum Torque Per Ampere
MVFB	Modified Variable Filter Bank
MWSN	Mobile Wireless Sensor Networks
NAN	Neighborhood Area Network
NAPT-PT	Network Address Port Translation-Pt

NARX	Nonlinear Auto Regressive Models With Exogenous Inputs
NAT	Network Address Translation
NAT-PT	Network Address Translation-Protocol Translator
NB	Naïve Bayes
NBS	Nash Bargaining Solution
NGD	Negative Group Delay
NIST	National Institute Of Standards and Technology
NLFB	Novel Lifting-Based Filter Bank
NLMS	Normalized Least Mean Square
NLP	Natural Language Processing
NMT	Neural Machine Translation
NoC	Network On Chip
NOMA	Non-Orthogonal Multiple Access
NPCR	Number of Pixel Change Rate
NPF	Next Day Forecast Power
N-P-K	Nitrogen-Phosphorous-Potassium
nrtPS	Non-Real-Time Polling Service
NS	Network Simulator
NTRU	n th Degree Truncated Polynomial Ring Units
NTT	Nippon Telegraph and Telephone
OA	Overall Accuracy
OFDM	Orthogonal Frequency Division Multiplexing
OMA	Orthogonal Multiple-Access
OMEGA	Observatoire Pour La Minéralogie, l'Eau, Les Glaces Et l'Activité
OOB	Out Of Band
OOK	On-Off Keying
OWL	Web Ontology Language
P	Predicted Value
P&O	Perturb And Observe
PaaS	Platform As A Service
PAPR	Peak To Average Power Ratio
PART	Projective Adaptive Resonance Theory
PC	Personal Computer
PCA	Principal Component Analysis
PCAP	Packet Capture
PCC	Predictive Current Control
PCC	Point of Common Coupling
PCU	Power Conditioning Unit
PDF	Probability Density Function
PDR	Packet Delivery Ratio
PE	Processing Element
PEC	Power Electronic Converters
PET	Positron Emission Tomography

PET	Prefixing Encapsulation and Translation
PFTIT2FLC	Passive Fault
PHE	Partial Homomorphic Encryption
PHEV	Plugin Electric Vehicle
Pi	Performance Index
P_i and P_j	Net Real Power Injection in Bus 'I' And 'J'
PID	Proportional Integral Derivative
PID	Proportional Integrator and Derivative
PIN	Personal Identification Number
PK	Public Key
PL	Physical Layer
PLC	Programmable Logic Controller
PLI	Power Loss Index
PLL	Phase-Locked Loop
PMDC	Permanent Magnet DC Generator
PMSM	Permanent Magnet Synchronous Machines
POP	Population
POS	Part-Of-Speech
PPDM	Privacy-Preserving Data Mining
PPG	Photoplethysmography
PPI	Protein-Protein Interaction
PPIM	Protein-Protein Interaction Database for Maize
PQ	Power Quality
PQE	Power Quality Events
PRBS	Pseudorandom Binary Sequence
PreLu	Parametric ReLU
PROMETHEE	Preference Ranking Organization Method for Enrichment of Evaluations
PRV	Predicted Variation
PSD	Power Spectral Density
PSO	Particle Swarm Optimization
PU	Primary Users
PV	Photo Voltaic
PV	Solar Photovoltaic Systems
PWM	Pulse Width Modulation
PY	Present Year
Q_i and Q_j	Net Reactive Power Injection in Bus 'I' And 'J'
QID	Quasi-Identifier
QMS	Quality Management System
QOp	Query Output
QoS	Quality of Service
QRC	Quasi Resonant Converter
QSS	Quality Six Sigma

R	Recall
R2L	Remote To Local
RAkEL	Random k LabELsets
RAM	Random Access Memory
RB	Rule Based
RBF	Radial Basis Function
RBR	Run-Based Reordering
RBS	Raiffa Bargaining Solution
RCNN	Region-Based Convolutional Neural Networks
RDF	Radiation Factor
ReLU	Rectified Linear Unit
RES	Renewable Energy Source
REST	Representational State Transfer
RF	Random Forest
RF	Radio Frequency
RFID	Radio Frequency Identification
RGB	Red Green Blue
RGC	Relative Growth Coefficient
R_{ij}	Resistance Between Bus 'I' And 'J'
RISC	Reduced Instruction Set Computer
RMSE	Root-Mean-Square Error
RNA	Ribonucleic Acid
RNN	Recurrent Neural Networks
ROC	Receiver Operating Curve
ROI	Region-Of-Interest
ROM	Read Only Memory
RPL	Routing Protocol for Low-Power and Lossy Networks
RPM	Rotations Per Minute
RPWM	Random Pulse Width Modulation
RRC	Root Raised Cosine
RSA	Rivest, Shamir, & Adleman
RSC	Rotor Side Converter
RSDCCF	Relevancy and Similarity Aware Drug Comment Classification Framework
RSPM	Respirable Suspended Particulate Matter
RTO	Real Time Optimizations
rtPS	Real-Time Polling Service
SA	Simulated Annealing
SaaS	Software As A Service
SAD	Smart Automation Device
SAIDI	System Average Interruption Duration Index
SAIFI	System Average Interruption Frequency Index
SAM	Standard Allowed Minute
SAP	Same Day Actual Power

SCA	Sine Cosine Algorithm
SCD	Spectral Correlation Density
SCM	Supply Chain Management
SDK	Software Development Kits
SDR	Software Defined Radio
SEGMSE	Segmental MSE
SEGSNR	Segmental SNR
SFP	Same Day Forecast Power
SG	Smart Grid
SGD	Smart Grid Device
SGD	The Saccharomyces Genome Database
SHA	Secure Hash Algorithm
SHARP	Stream-Based Holistic Analytics and Reasoning in Parallel
SHE	Somewhat Homomorphic Encryption
SHR	Self-Healing Ring
SIC	Successive Interference Cancellation
SIIT	Stateless Ip/Icmp Translation Algorithm
SK	Secret Key
SLA	Service Level Architecture
SLFN	Single Layer Feed-Forward Neural Network (a neural network algorithm used for classification)
SLIC	Segmentation-Based Lossless Image Coding
SLM	Selective Mapping
SM	Synchronous Motor
SMLR	Sparse Multinomial Logistic Regression
SNR	Signal to Noise Ratio
SOA	Service-Oriented Architecture
SOAF	Second Order Adaptive Notch Filter
SoC	System On Chip
SOC	State Of Charge
SOCKS	Socket Secure
SOIW	Subjective and Objective Integrated Weightage
SOP	Secrecy Outage Probability
SOQL	Salesforce Object Query Language
SP	Set Point
SPDs	Surge Protection Devices
SPWM	Single Pulse With Modulation
SQLMR	Structure Query Language Map Reduce
SRF	Sobol Random Function
SRM	Switched Reluctance Motor
SRP	Sobol Random Permutation
SSS	Shunt Series Switched
St	Settling Time
ST	S-Transform

STATCOM	Static Synchronous Compensator
STC	Standard Test Conditions
SU	Secondary User
SVC	Static Var Compensator
SVM	Support Vector Machine
SVR	Support Vector Regression
SWA	Sliding Window Algorithm
T	Process Time Constant
t	Equivalent Time Constant
T1FLC	Type 1 Fuzzy Logic Controller
TANGEDGO	Tamil Nadu Generation Distribution Corporation
TCF	Temperature Change Factor
TCL -	Tool Command Language
TCP/IP	Transmission Control Protocol/Internet Protocol
T_d	Derivative Time
TER	Translation Error Rate
THD	Total Harmonic Distortion
T_i	Integral Time
TN	True Negative
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TP	True Positive
TPA	Third Party Auditor
TPL	Total Real Power Losses Without DG
TPL_{DG}	Total Real Power Losses With DG
TPR	True Positive Rate
TRESTBPS	Resting Blood Pressure
TRNSYS	Transient System Simulation Tool
TRT	Transport Relay Translator
TSVM	Transductive Support Vector Machines
T_u	Time Unit
TV	Term Variance
TV_{sum}	Sum of the Total Variation
U2R	User to Root
UACI	Unified Average Changing Intensity
UC	Ultracapacitor
UFMC	Universal Filtered Multi-Carrier
UFOPTD	Unstable First Order Plus Time Delay
UG	Utility Grid
UGS	Unsolicited Grant Service
UHF	Ultrahigh Frequency
UIN	Unique Identification Number
UPS	Uninterrupted Power Supply
URL	Uniform Resource Locators
USN	Ultrasonic Sensor Node

USRP	Universal Software Radio Peripheral
UWB	Ultra-Wide-Band Antenna
UX	User Experience
V4GT	Ipv4-Enabled Gateway Translator
V6GT	Ipv6-Enabled Gateway Translator
VCO	Voltage Controlled Oscillator
VDR	Voltage Double Rectifier
VFFB	Voltage-Fed Full Bridge
V_i and V_j	Voltage at Bus 'I' And 'J'
VIHC	Variable Length Input Huffman Coding
VIN	Vehicle Identification Number
VLC	Visible Light Communication
VLIW	Very Long Instruction Word
VoI	Value of Information
VSI	Voltage Source Inverter
VSNL	Videsh Sanchar Nigam Limited
w/v	Weight/Volume
WAN	Wide Area Network
WD	Weibull Distribution
WECS	Wind Energy Conversion System
WEV	Word Embedding Visualization
WIFI	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access
WITH ST	With Stop Words
WITHOUT ST	Without Stop Words
WOA	Whale Optimization Algorithm
WoTs	Web of Things
WPM	Weighted Product Method
WSD	Word Sense Disambiguation
WSM	Weighted Sum Method
WSN	Wireless Sensors Networks
WT	Wind Turbine
WTG	Wind Turbine Generator
ZCS	Zero Current Switch
ZF NOMA	Zero Forcing Non-orthogonal Multiple Access
ZFBF	Zero Forcing Beamforming
ZVS	Zero Voltage Switch
δ_i and δ_j	Angle at Bus 'I' And 'J'
%OV	Percentage Overshoot
$ I_{ij} ^{\max}$	Maximum Limit of Current in Bus 'I' And 'J'
$ V_i ^{\min}$ and $ V_i ^{\max}$	Minimum and Maximum Limit of Voltage in Bus 'I'
2DOF	Two Degrees of Freedom
6LoWPAN	Ipv6 Low Power Wireless Personal Area Network

Chapter 1

Fractional-Order PID Controller Optimized by SCA for Solar System



Raj Kumar Sahu, Binod Shaw, and Jyoti Ranjan Nayak

Abstract In this paper, the maximum power point tracking (MPPT) technique is enhanced by implementing fractional-order proportional-integral derivative (FOPID) to improve the output of DC-DC boost converter of solar system and justified by competing alongside perturb and observe (P&O) and PID-based MPPT technique. The gain variables of FOPID and PID controller highly influenced the performance of the system. Sine Cosine Algorithm (SCA) is a novel technique to explore the best couple of gain factors of both PID and FOPID controllers to provide appropriate gate pulse to converter to improve performance of the output response. In this paper, the performance of the photo-voltaic (PV) system is enhanced by conceding the oscillation, time response, settling time and maximum values of voltage, current and power of the system by adopting FOPID controller.

Keywords Perturb and observe (P&O) · Proportional-integral-derivative (PID) controller · Fractional-order PID (FOPID) · Sine Cosine Algorithm (SCA)

Abbreviations

FOPID	Fractional-order PID
MPPT	Maximum power point tracking
P & O	Perturb and observe
PID	Proportional-integral-derivative
PV	Photo-voltaic
SCA	Sine Cosine Algorithm

R. K. Sahu (✉) · B. Shaw · J. R. Nayak
National Institute of Technology, Raipur, Chhattisgarh, India

1.1 Introduction

Most of the power generations are from conventional source of energy such as thermal, nuclear, etc. Due to increase in population, lack of natural resources (coal, petroleum, gas, etc.) and growing industries, the renewable energy sources contribute to meet the demand. Solar PV is one of the most imperative approaches among the renewable energy sources. The solar system dominates over other renewable sources are as follows: it is eco-friendly, easy to maintain, easy to install, and noise-free and it has impressive life span. The solar PV system depends on solar energy, its intensity and irradiation. Due to its dependency, its performance is non-linear. Hence to improve its performance, MPPT technique is used to improve the output response of DC-DC boost converter and it generates the maximum power by adjusting the load resistance appropriately in variable atmospheric condition.

1.2 Literature Survey

Numerous scientists have executed different procedures to improve the productivity of electricity of the solar-based PV cell by upgrading the MPPT systems from most recent couple of decades. ESRAM et al. [1] have contributed 19 diverse MPPT procedures and gave a reasonable elucidation to analysts to receive pertinent strategies. Perturb & Observation (P&O) uses variable step size with the progression estimate is self-adjusted and analysis alongside the traditional technique in [2]. Ishaque and Salam have contributed a concise writing to structure MPPT by implemented soft computing techniques amid of variable irradiance in [3]. Application of soft computing technique-implemented MPPT algorithm for variable irradiance is well explained by the variable Cuckoo Search (CS). MPPT calculation is justified by analysis with traditional P&O and Particles Swarm Optimization (PSO) MPPT techniques in three different contextual analyses and is portrayed in [4]. A brief study of various research papers on MPPT technique is portrayed in [5, 6]. An optimized P&O technique is implemented to improve the performance of PV system in [7]. Adaptive fuzzy-PI controller is implemented as MPPT and the role of climate change on PV module is well established in [7, 8], respectively. Various soft computing techniques and optimization techniques are adopted to enhance the performance of MPPT of PV module in [9–11]. In this paper, SCA algorithms are implemented in FOPID based MPPT technique to justify over P&O and PID controller with enhancement in the output performance of PV system by supplying proper duty cycle to the switch of boost converter. MATLAB/Simulink platform executed the proposed design of PV system.

1.3 Proposed PV System

The proposed PV system with FOPID controller is depicted in Fig. 1.1, a PV module SunPower SPR 315E WHT D with one parallel and one series string with 96 cells per module. The specifications of PV module parameters are maximum power, open circuit voltage, short circuit current, voltage at maximum power point (MPP) and current at maximum power point are 315.072 W, 6.6 V, 6.14 A and 5.76 A, respectively. The block diagram of proposed isolated solar system is portrayed in Fig. 1.2. The regulated duty cycle of the converter improves the performance of the solar system.

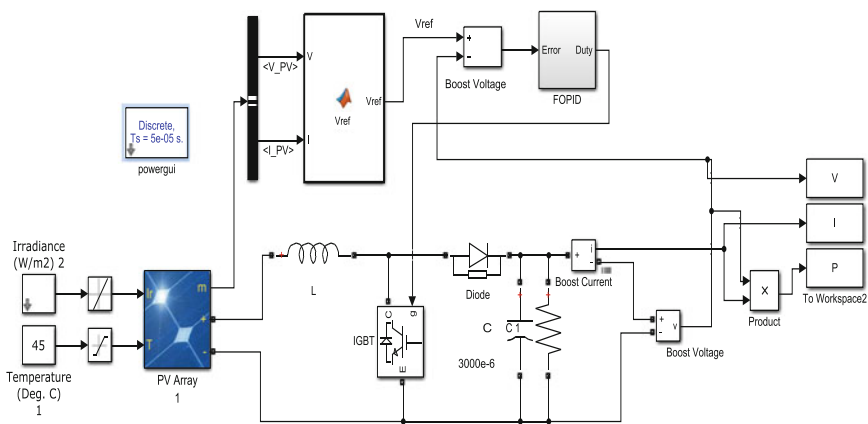


Fig. 1.1 Simulink model of PV system with FOPID controller

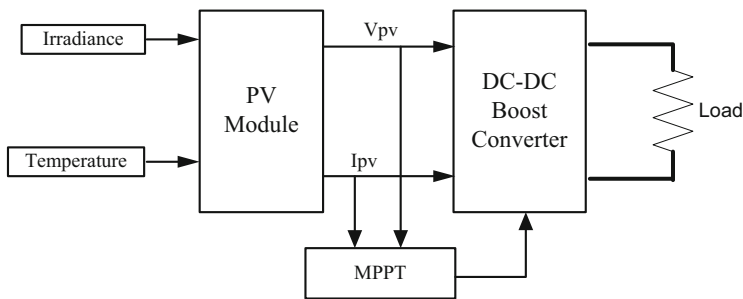


Fig. 1.2 Block diagram of PV system

1.3.1 PV Module

PV cells are represented by the equivalent current source, diode, series resistor and parallel resistors as portrayed in Fig. 1.3. The ideal solar system may be explained through Eqs. (1.1, 1.2, 1.3 and 1.4):

$$I_{rr} = I_{scr} e^{(qV_{oc}/KN_sAT_{rk})-1}, \quad (1.1)$$

$$I_d = I_{rr}(T_{ak}/T_{rk})^3 e^{[(E_gK/KA)(1/T_{rk}-1/T_{ak})]}, \quad (1.2)$$

$$I_{PH} = I_{scr} + (K_i(T_{ak} - T_{rk}))S/1000, \quad (1.3)$$

$$I_o = N_p I_{PH} - N_p I_d \left\{ e^{(q/N_s AKT_{ak})(V_o + I_o R_s)} - 1 \right\}, \quad (1.4)$$

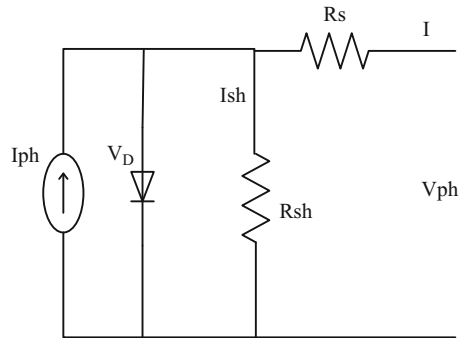
where I_o = PV module current, V_o = PV module voltage, T_{rk} = reference temperature in Kelvin, T_{ak} = operating temperature in Kelvin, S = irradiance W/m^2 , q = charge of electron, 1.6×10^{-19} C, A = ideality factor, 1.3, K = Boltzman constant, E_g = band gap, I_{scr} = S.C current, N_s = cells connected in series, N_p = cells connected in parallel, K_i = S.C temperature coefficient, R_s = series resistance and I_{Ph} = light-generated current.

The output of a solar cell module is about 12 V which is generally low for an application. In order to increase the output voltage, a number of modules are connected in series/parallel. To maintain a constant output voltage and to improve voltage profile, a boost converter is portrayed in Fig. 1.4. Voltage profile is improved by controlling the gate pulse of the converter. The pulse of a converter is achieved by Pulse width Modulation (PWM) technique. The output of the converter may be characterized in Eq. (1.5):

$$V_{out} = \frac{1}{1-D} V_{in}, \quad (1.5)$$

where D is the duty cycle of the converter and is characterized in Eq. (1.6):

Fig. 1.3 Equivalent circuit of solar cell



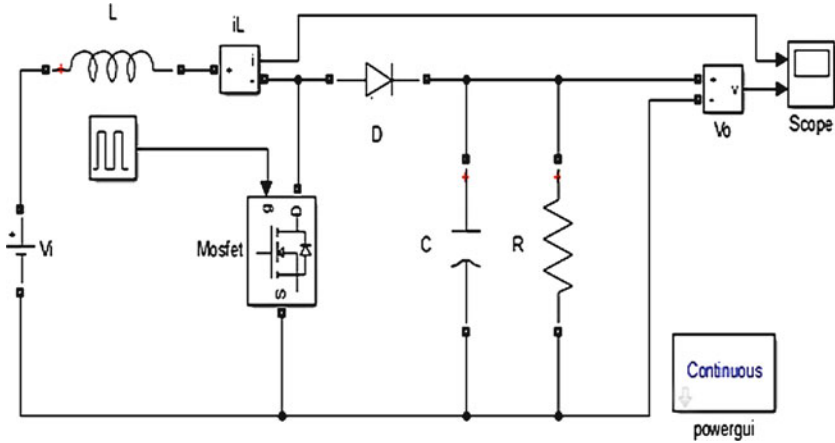


Fig. 1.4 Boost converter

$$D = \frac{t_{on}}{t_{on} + t_{off}}. \tag{1.6}$$

On time and off time of the switch are expressed in Eqs. (1.7) and (1.8), respectively, by conceding switching period (T_s):

$$t_{on} = DT_s, \tag{1.7}$$

$$t_{off} = (1 - D)T_s. \tag{1.8}$$

The parameter of boost converter inductance (L), capacitance (C) and load (R) are $1 \mu\text{H}$, $3000 \mu\text{F}$, and 24Ω , respectively.

1.3.2 FOPID ($PI^\lambda D^\mu$) Controller

FOPID is a new methodology suggested as the fractional order. Here, two additional parameters, fractional-order integrator (λ) and differentiator (μ), are used as compared with PID controller. The transfer function of the $PI^\lambda D^\mu$ controller is characterized in Eq. (1.9):

$$G_C(s) = K_P + \frac{K_I}{s^\lambda} + s^\mu K_D. \tag{1.9}$$

It has matchless control quality compared with PID controller to keep up sound of the proposed system due to fraction order. FOPID controller structure is depicted in Fig. 1.5. Error is the difference of reference voltage (V_{ref}) calculated P&O-based algorithm shown in Fig. 1.6 and boost output voltage.

Fig. 1.5 FOPID controller structure

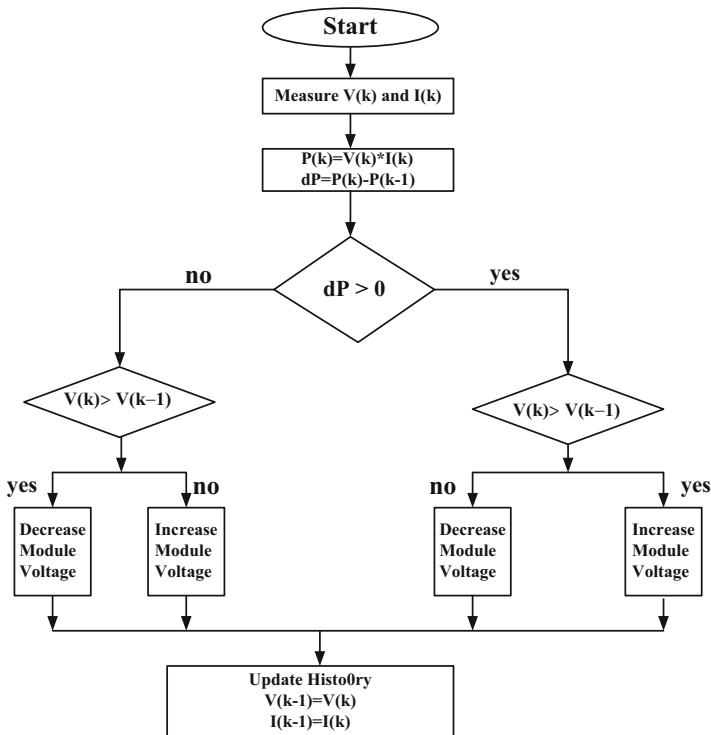
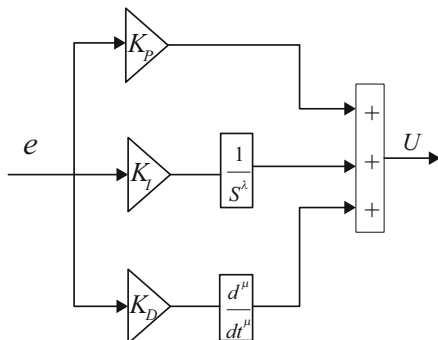


Fig. 1.6 P&O-based V_{ref} calculation algorithm

1.4 Sine Cosine Algorithm (SCA)

In this paper, the PID- and FOPID-based MPPT controller schemes are designed to contribute the appropriate generated duty cycle (gate pulse) of the converter. In this paper, a novel technique proposed by Mirjalili [12] is implemented to tune the controllers. The mathematical model of SCA is established using sine and cosine

functions. Various random values are included to counterbalance the exploration and exploitation of search space. The steps followed for SCA are described as:

1. Prepare the population matrix of size $[P \times D]$, i.e. $[X]P \times D$, where P and D are the population numbers and numbers of gain factor. For this problem, the specific limit of the designed variables is 0.01–2.
2. Evaluate the fitness for all populations to determine the destination point for that iteration, i.e. P_i , and the position of the population is updated using (1.10) and (1.11):

If $r_4 < 0.5$

$$x_i^{t+1} = x_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t|, \quad (1.10)$$

otherwise

$$x_i^{t+1} = x_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t|, \quad (1.11)$$

where r_1 is the decaying function and may be expressed as $r_1 = a - (a \times i) / \text{itermax}$

where ‘ a ’ is a constant and i is the current iteration. r_2 is the random value between specific limit of $0-2\pi$ and r_3 is the random number between 0 and 1.

3. Update the X matrix by comparing the fitness value. The better fitness population will exist for the next generation, and repeat the steps from 2 to 4 until the end criteria are reached. In this issue maximum repetition is used as the end criterion.

In this paper, SCA algorithms are applied to tune the gain parameter of PID and FOPID controllers to improve the output voltage and current of the PV module.

1.5 Results and Discussion

SCA calculation is executed for 50 cycles with 50 populations to determine the means to find the ideal gain parameters of both PID and FOPID controllers. The goal of the calculation is to chase the parameters inside a predefined limit as depicted in Eqs. (1.12) and (1.13):

$$0.001 \leq K_p, K_I \quad \text{and} \quad K_D \leq 2, \quad (1.12)$$

$$0.001 \leq \mu \quad \text{and} \quad \lambda \leq 2 \quad (1.13)$$

The optimal values of PID controller, i.e. K_p , K_I and K_D , are 0.0711, 0.9079 and 1.1260, respectively. The optimal values of FOPID controller, that is, K_p , K_I , K_D , μ and λ are 0.5211, 0.8594, 0.5958, 0.2393 and 1.4131, respectively. The execution of PV modules by surrendering power, voltage and current are depicted in Figs. 1.7, 1.8 and 1.9, respectively. $I-V$ and $P-V$ characteristics of PV module are shown in Fig. 1.10.

Fig. 1.7 Power versus time graph

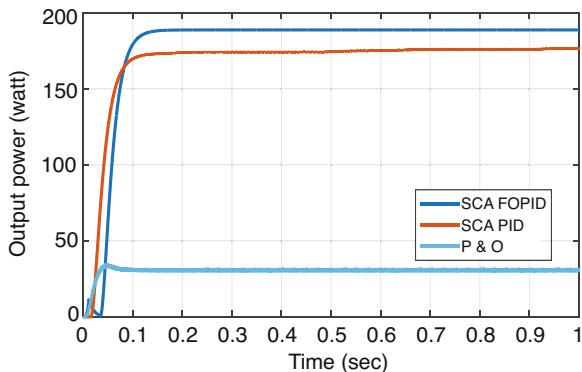


Fig. 1.8 Voltage versus time graph

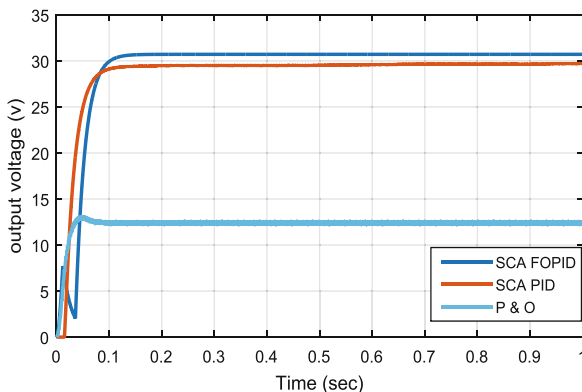
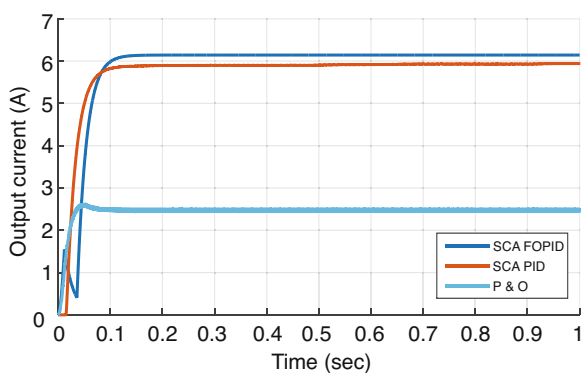


Fig. 1.9 Current vs time graph



In this proposed PV system, SunPower SPR-315E-WHT-D PV module with one parallel and one series string are taken. The $I-V$ and $P-V$ characteristics of PV module are shown in Fig. 1.10. The performance parameters of the output of PV module are tabulated in Table 1.1 to provide a fair supremacy of FOPID-based MPPT controller over P&O and PID technique.

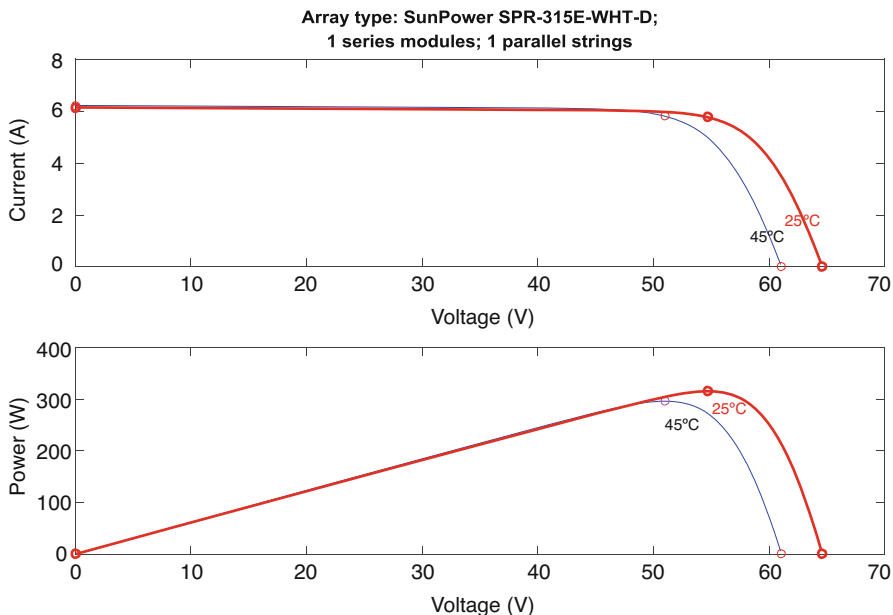


Fig. 1.10 $I-V$ and $P-V$ characteristics of PV module

Table 1.1 Performance response of output

MPPT techniques	1. Maximum voltage	2. Maximum current	3. Maximum power
P&O	12.63	2.526	31.92
PID	29.5	5.9	173.8
FOPID	30.69	6.142	188.6

1.6 Conclusion

The persistence of this work is to plan a PV system and to improve the productivity of the system by applying FOPID controller. The FOPID and PID controller parameter enhanced by SCA is validated as a better controller over P&O-based controller of the solar system. Temperature, irradiance and load are the imperative factors which influence the current, voltage and power of the solar cell. This FOPID, PID and P&O MPPT controller are performed by diverging the irradiance ($1000-700 \text{ W/m}^2$), with constant temperature (45°C) and load (24Ω). The difference signal is assessed by the difference of reference voltage and actual process voltage to accomplish regulated gate pulse (duty cycle) of the converter. The output voltage is massively affected by duty cycle of the gate pulse. The SCA streamlined FOPID-based MPPT algorithm is validated over P&O and PID method to achieve maximum power, current and voltage.

References

1. ESRAM T, CHAPMAN PL (2007) Comparison of photovoltaic array maximum power point tracking techniques. *IEEE Trans Energy Convers* 22(2):439–449
2. AL-DIAB A, SOURKOUNIS C (2010) Variable step size P&O MPPT algorithm for PV systems. *Proc Int Conf Optim Electr Electron Equip OPTIM*:1097–1102
3. ISHAQUE K, SALAM Z (2013) A review of maximum power point tracking techniques of PV system for uniform insolation and partial shading condition. *Renew Sust Energ Rev* 19:475–488
4. AHMED J, SALAM Z (2014) A maximum power point tracking (MPPT) for PV system using Cuckoo Search with partial shading capability. *Appl Energy* 119:118–130
5. KAMARZAMAN NA, TAN CW (2014) A comprehensive review of maximum power point tracking algorithms for photovoltaic systems. *Renew Sust Energ Rev* 37:585–598
6. REZA A, HASSAN M, JAMASB S (2013) Classification and comparison of maximum power point tracking techniques for photovoltaic system: a review. *Renew Sust Energ Rev* 19:433–443
7. FEMIA N, PETRONE G, SPAGNUOLO G, VITELLI M (2005) Optimization of perturb and observe maximum power point tracking method. *IEEE Trans Power Electron* 20(4):963–973
8. HOFFMANN S, KOEHL M (2013) Investigation on the impact of macro- and micro-climate on the potential induced degradation, pp 1822–1825
9. CHENG PC, PENG BR, LIU Y-H, CHENG YS, HUANG JW (2015) Optimization of a fuzzy-logic-control-based MPPT algorithm using the particle swarm optimization technique. *Energies* 8 (6):5338–5360
10. NAGARAJAN R, YUVARAJ R, HEMALATHA V, LOGAPRIYA S, MEKALA A, PRIYANGA S (2017) Implementation of PV – based boost converter using PI controller with PSO algorithm. *Int J Eng Comput Sci* 6(3):20479–20484
11. NEÇAİBIA A, LADACI S, CHAREF A, LOISEAU JJ (2015) Fractional order extremum seeking approach for maximum power point tracking of photovoltaic panels. *Front Energy* 9(1):43–53
12. MIRJALILI S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133

Chapter 2

LVRT Capability Improvement in a Grid-Connected DFIG Wind Turbine System Using Neural Network-Based Dynamic Voltage Restorer



Arun Kumar Puliyadi Kubendran and L. Ashok Kumar

Abstract The wind power plant is one of the fastest growing electrical power sources. The integration of wind turbine into the power grid originates various power quality problems. Low-voltage ride through (LVRT) capability improvement is impregnating exigency in recent power quality issue, which is acquainted with various renewable power generation resources like solar PV and wind power plant. Dynamic voltage restorer (DVR) is a custom power device (CPD), which is connected in series with the electrical system, and it is a contemporary and efficient CPD expended in the distribution system to curb the power quality problems. In this paper, LVRT capability of doubly fed induction generator (DFIG) wind turbine system connected to a power grid is enhanced by means of DVR. LVRT capability is improved by the introduction of neural network controller. This controller fulfils the various grid code requirements. Thus, the performance of the DVR becomes far superior by the robust control technique. The simulation results were compared with conventional PI controller and the proposed artificial neural network (ANN) controller. It is proved that the proposed system raises the reliability of grid-connected DFIG wind turbine system.

Keywords Wind farm · Low-voltage ride through · DVR · Custom power device · DFIG · Neural network · Power quality

A. K. Puliyadi Kubendran (✉)
Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India

L. Ashok Kumar
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

Abbreviation

ANN	artificial neural network
CPD	custom power device
DFIG	doubly fed induction generator
DVR	dynamic voltage restorer
EMF	electromotive force
FNN	feed-forward neural network
GSC	grid side converter
LVRT	low-voltage ride through
PQ	power quality
RSC	rotor side converter
STATCOM	static synchronous compensator
SVC	static VAR compensator
WECS	wind energy conversion system
WT	wind turbine
WTG	wind turbine generator

2.1 Introduction

At present, wind energy is one of the most sustainable energy resources. The remarkable development of large-scale grid-coupled wind farms has turned into a necessary part in the entire energy assembly [1]. The part of wind power plant with respect to toughness of electricity grid causes issues to system operators, such as voltage deviations, grid voltage destabilisation and grid uncertainty [2, 3]. Recently, wind farms are not reflected as a modest energy source but functioned as power plants and able to deliver reactive power, keep on connected during system faults, and adapt their control subject on requirements of the system, so that numerous novel requirements for wind turbines are refined by system operators and LVRT capability is the most significant one [4]. Numerous schemes are considered to enhance LVRT capability of wind turbine generators (WTG) in different types [5]. The most known solution for DFIG under the faulty condition is to fix a crowbar on the rotor circuit [6], so that for the duration of voltage sag DFIG can perform like a conventional induction generator and yield reactive power from grid. The main argument over this solution creates serious stability problem. The LVRT capability of DFIG can be improved by reactive power injection device such as static VAR compensator (SVC) [7, 8] and static synchronous compensator (STATCOM) [9, 10, 11], but the grid code demand cannot be met. In this paper, a DVR is planned to increase the LVRT capability of WTGs. The chief benefit of DVR is its capability to shoot up a controllable reactive current independently on the grid voltage, thus compensating the current [12, 13]. Here a different type of fault is examined as this fault produces a voltage sag in the grid-connected wind energy

conversion system (WECS). LVRT capability of DVR with DFIG-connected WECS is implemented with the help of MATLAB simulations. The results demonstrate the performance of wind power plant under various fault conditions. In Sect. 2.2 DFIG wind turbine system is discussed. The LVRT capability is evaluated in Sect. 2.3. Section 2.4 deals with DVR modelling. The various control strategies of DVR are explained in Sect. 2.5. The compensation of DVR during LVRT fault is represented in Sect. 2.6 and followed by conclusion.

2.2 Wind Turbine System

An electric grid-connected back-to-back converter with active and reactive power wind turbine fitted with the DFIG is considered in this paper [14–17]. The stator is directly coupled to the system whereas the rotor is linked through a rotor side converter (RSC), the dc-link, a grid side converter (GSC) and a transformer to the grid. The need of the GSC control is to continue the dc-link capacitor voltage in a set value irrespective of the magnitude and also the direction of the rotor power to certify a converter operation with unity power factor. The DFIG system has a smaller power electronic converter and a lower investment cost as well as less converter losses. A problem with this system is its inability to ride-through severe voltage dips. Newly, manufacturers of DFIG systems provide solutions for management of all dips indicated in the grid codes. The wind turbine mechanical power P_t is converted into the electric power P_g by DFIG which is then imparted to the grid. The DFIG model with standard operation and LVRT operation facility is developed, and analysis is performed using simulation. The main objective is to choose the controller parameters for this system.

2.3 Low-Voltage Ride Through (LVRT)

One of the significant requirements of grid code is low-voltage ride through (LVRT) capability. This requirement usually differs from one country to another since it depends on the specific characteristics of each power system and the protection employed. In general, wind turbines (WTs), under LVRT requirement, are basically forced to remain connected to the grid during voltage interruptions or voltage dips with remaining voltage of almost 15% of the nominal voltage levels. Generally, the whole LVRT process for DFIG-based WECS can be divided into three continuous periods according to the LVRT grid codes, the initial period after the grid fault occurs (first period), the low voltage-sustaining period (second period) and the grid voltage recovery period after fault clearance (the third period). To achieve complete LVRT, the control objectives for DFIG-based WECS in each period are different. In the initial period after a grid fault occurs, a large electromotive force (EMF) is induced in the rotor circuit of the DFIG by the transient decay component of the stator flux, which results from the stator voltage dip. To ensure the safe operation of

the rotor-side converter (RSC), the control objective for DFIG-based WECS in this period is to restrain the rotor surge current caused by the large EMF. When the transient stator flux is almost damped out, it enters the low voltage-sustaining period. The control objective for DFIG-based WECS in this period is to support the grid voltage by injecting reactive current into the network. Grid contains many distributed generators when subjected to low voltage disconnect produces chain reaction. The voltage in the grid drops further as any of the generator is removed due to voltage drop. This leads to cascading failure. The difference in active power is managed by LVRT requirement. The constrain behind LVRT is to make the wind power generation system connected to the grid overcoming voltage dips and also to deliver the reactive power to the grid to maintain grid voltage. LVRT capability is plotted between voltage to the time characteristics and the slope of recovery depends on the strength of interconnection and reactive power support. The amount of reactive power added to the grid depends on the percentage of grid voltage reduction during the dip, the system rated current, and the reactive power delivered to the grid during the voltage sag. The LVRT curve shows the immunity of wind power station toward fault in the grid. The LVRT requirements also account fast active and reactive power restoration.

2.4 Dynamic Voltage Restorer (DVR)

A DVR comprises of a GTO or IGBT, capacitor bank which acts as energy storage device and injection transformers; on whole it is called as a solid-state power electronic switching device [18, 19]. The function of DVR is to produce a control voltage with an injecting transformer by make use of a forced commutated converter connected in series with bus voltage. DVR is completely suited to safeguard sensitive loads from voltage dips and swells. DVR has three modes of operation. If no voltage sag is occurred in the system, the DVR operates in standby mode. DVR function in injection mode whenever voltage sag occurs. If the current exceeds the permissible level at the load side the DVR operates in the protection mode and gets isolated from the system. The DVR stay connected to the grid provided the line breaker does not trip and it operates individually to the type of fault or any event that occurs in the system.

2.5 Control Strategy of DVR

2.5.1 Control Based on PI Controller

The proportional plus integral controller is called as PI controller. It has two gain constants which include the proportional gain k_p and integral gain k_i . In this manuscript, the value of proportional controller is considered as 100 and the value of integral controller is considered as 0.1. When the voltage sag or swell has

occurred, it is sensed and proper gain setting is made in the controller and based on the control the PWM generator generates the required pulses to the inverter. AC voltage is received from the inverter and injected using injection transformer, which is fed to the system in par with the bus voltage.

2.5.2 Control Based on ANN Controller

An elementary feed-forward neural network (FNN) architecture is considered in this manuscript, which consists of two layers encompassing the input layer and output layer [20, 21]. This type of network can be connected in cascade to create a multilayer FNN. Even though the FNN has no explicit feedback connection, the output values are often compared with the desired output values, provided by a ‘supervisor’. The error between the outputs of the neural network and the desired values can be employed for adapting the network’s weights. By modifying the weights, the error decreases. A set of input and output patterns called a training set is required for this learning mode. The input layer enables the connection to the hidden layer to transfer the input signal. Process of learning happens at the input layer which is followed by the output layer. The linear transfer function is enabled by output layer. The three inputs of ANN are the three phase voltage errors and the switching functions are generated by the three-output layer which delivered to the inverter legs. The ANN controller output is given as reference value for the PWM generator. It is compared with the existing system signal for injecting the required voltage to the system. During LVRT fault, the voltage sag occurs, the magnitude of the output of ANN controller is more and accordingly the PWM circuit generates high output. When a swell condition arises, the magnitude of ANN controller is lesser than the system signal then the PWM produces low output. Based on the weights and bias functions the learning of the neurons. If the learning rate is larger, the process of learning would be quick. The condition of instability occurs when the learning rate is higher which also increases errors.

2.6 Simulation Results and Discussions

LVRT capability improvement in a grid-connected WECS using DVR with PI controllers for different fault condition is simulated separately and simultaneously in the MATLAB, and the results including the THD analysis for injected voltage to overcome LVRT fault are presented in the Figs. 2.1, 2.2 and 2.3. Similarly, the same procedure is continued with FNN controller and the simulation results are represented from Figs. 2.4, 2.5 and 2.6.

From the execution, LVRT capability improvement using DVR with the help of PI and ANN controller injects voltage to the grid during various fault conditions and the FFT analyses of both controllers are represented and it is perceived that the total

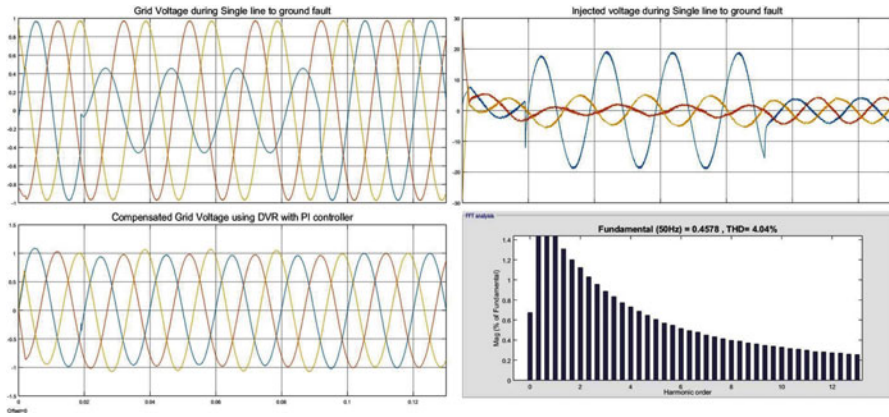


Fig. 2.1 Output waveforms of DVR using PI controller for single line to ground fault

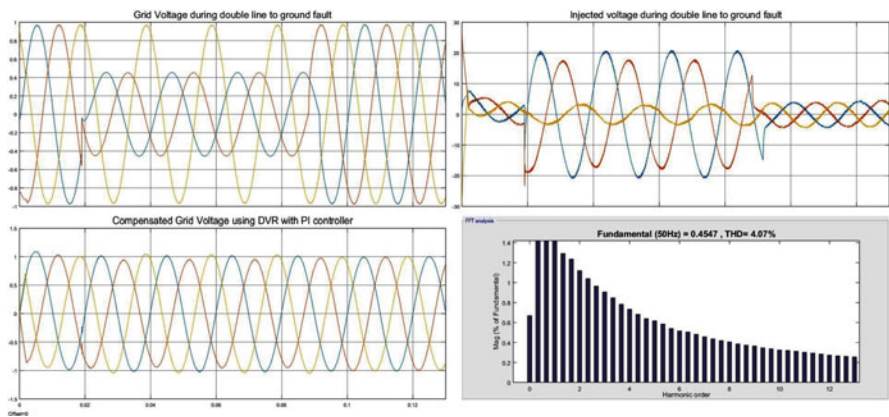


Fig. 2.2 Output waveforms of DVR using PI controller for double line to ground fault

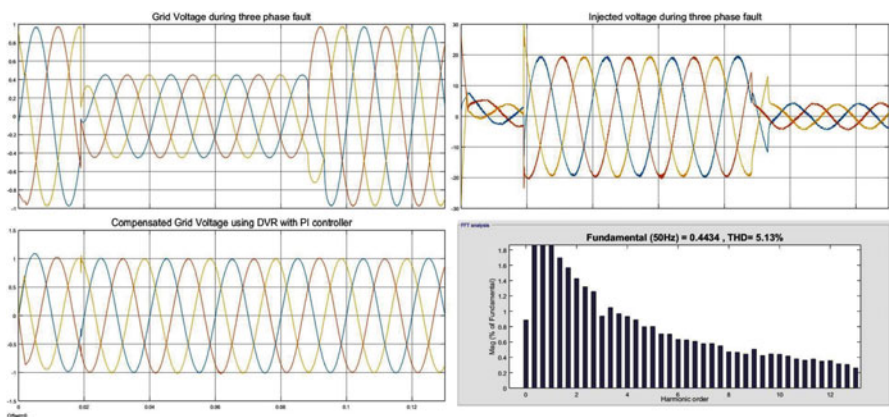


Fig. 2.3 Output waveforms of DVR using PI controller for three-phase fault

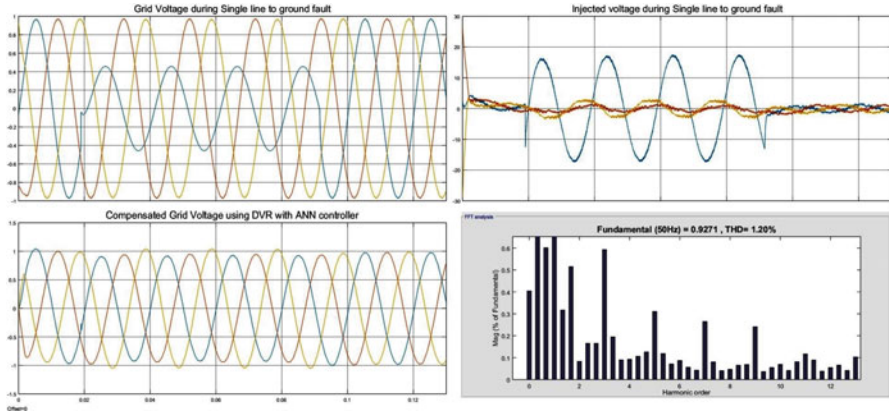


Fig. 2.4 Output waveforms of DVR using ANN controller for single line to ground fault

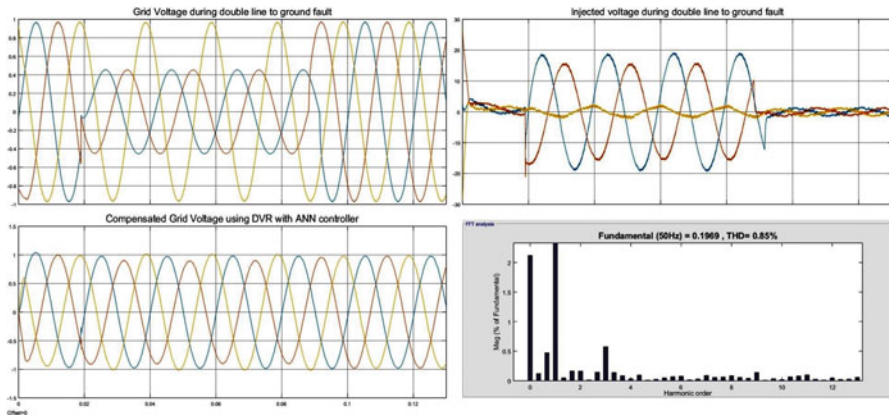


Fig. 2.5 Output waveforms of DVR using ANN controller for double line to ground fault

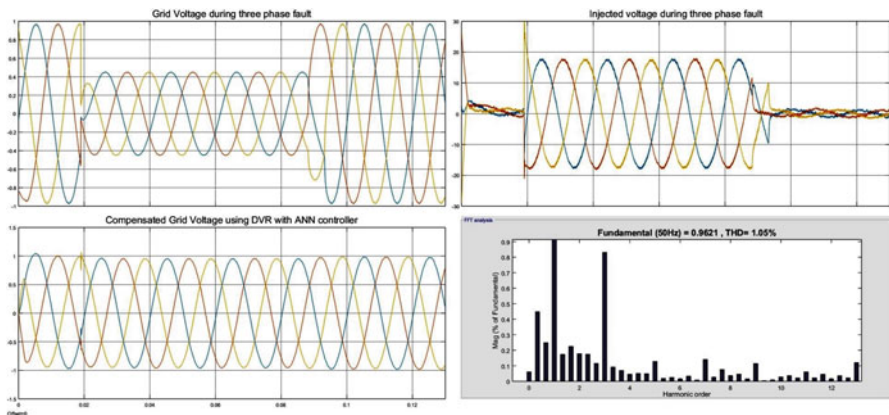


Fig. 2.6 Output waveforms of DVR using ANN controller for three-phase fault

Table 2.1 THD value comparison of LVRT fault injected voltage

Fault type	THD of PI controller (in %)	THD of ANN controller (in %)
Single line to ground	4.04	1.20
Double line to ground	4.07	0.85
Three-phase	5.13	1.05

harmonic distortion (THD) of ANN controller is lesser and in admissible range according to the IEEE standard 519–1992.

From Table 2.1, it is viewed that the THD of ANN controller shows better performance than PI controller with reduced THD percentage. It is also transparent that with the increase in severity of the fault the values of NN goes on decreasing while PI being constant. From the transient characteristic analysis, it is seen that NN controller injects voltage 0.002 s prior to PI controller.

2.7 Conclusion

The growth of large-scale grid-connected wind farms is increasing very rapidly during the last years. The power system stability is often disturbed due to the protective disconnection of a large amount of wind power during voltage sags. This condition is eradicated by improving the LVRT capability of wind turbines. As per the new grid code requirement, the LVRT requirement states that wind turbines should remain connected to the grid when a grid voltage sag is sensed even at faults due to short-circuit. In this paper, DVR has been proposed to provide an effective LVRT scheme for WTGs. This device produces an additional support during the voltage dip and keeps the WTG terminal voltage normally. Finally, a detail simulation model is built in MATLAB and from the simulation results it is clear that DVR performs efficiently with the ANN controller. During the conditions of voltage sag, the DVR injects voltage with proper response time, i.e. 0.002 s prior to PI controller. Albeit the severity of the faults increases the harmonic distortion values the injected voltage also gets reduced only in the ANN controller. Thus, in the grid-connected wind energy conversion system, the performance of the DVR is superior with ANN controller. The simulation results verify the feasibility and usefulness of the proposed device and control strategies.

References

1. Chai C, Lee WJ, P F et al (2005) System impact study for the interconnection of wind generation and utility system. *IEEE Trans Ind Appl* 41(1):163–168
2. Kubendran AKP, Loganathan AK (2017) Detection and classification of complex power quality disturbances using S-transform amplitude matrix-based decision tree for different noise levels. *Int Trans Electr Energy Syst* 27(4):1–12. <https://doi.org/10.1002/etep.2286>

3. Kumar PKA, Vijayalakshmi VJ, Karpagam J, Hemapriya CK (2016) Classification of power quality events using support vector machine and S-Transform. 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, pp 279–284. <https://doi.org/10.1109/IC3I.2016.7917975>
4. Case C (2006) Connecting wind farms to the grid-what you need to know. Vancouver
5. Xing Z, Longyun Z, Shuying Y et al (2008) Low voltage ride-through technologies in wind turbine generation. *Proc CSU-EPSA* 20(2):1–8
6. Ottersten R, Petersson A, Pietilainen K (2004) Voltage sag response of PWM rectifiers for variable-speed wind turbines. Nordic workshop on power and industrial electronics, Chalmers University of Technology
7. Chompoo-inwai C, Yingvivananpong C, Methaprayoon K et al (2005) Reactive compensation techniques to improve the ride-through capability of wind turbine during disturbance. *IEEE Trans Ind Appl* 41(3):666–672
8. Muller S, Deicke M, De Doncker RW (2002) Doubly fed induction generator systems for wind turbine. *IEEE Ind Appl Mag* 3:26–33
9. Pena R, Clare JC, Asher GM (1996) Doubly fed induction generator using back-to-back PWM converts and its application to variable speed wind-energy generation. *IEE Proc Electr Power Appl* 143:231–241
10. Arunkumar PK, Kannan SM, Selvalakshmi I (2016) Low voltage ride through capability improvement in a grid connected wind energy conversion system using STATCOM. International Conference on Energy Efficient Technologies for Sustainability (ICEETS), Nagercoil, pp 603–608
11. Solar PV and wind energy conversion systems: an introduction to theory, modeling with MATLAB/SIMULINK, and the role of soft computing techniques, S Sumathi, LA Kumar, P Surekha, Springer, Green Energy and Technology, <https://doi.org/10.1007/978-3-319-14941-7>, ISBN: 9783319149400, 9783319149417 (online)
12. Tapia A, Tapia G, Ostolaza JX, Saenz JR (2003) Modeling and control of a wind turbine driven doubly fed induction generator. *IEEE Trans Energy Conver* 18:194–204
13. Lei Y, Mullane A, Lightbody G, Yacamini R (2006) Modeling of the wind turbine with a doubly fed induction generator for grid integration studies. *IEEE Trans Energy Conver* 21(1):257–264
14. Akhmatov V, Krudsen H (1999) Modelling of windmill induction generator in dynamic simulation programs. In: Proceedings of IEEE International conference on power technology, Budapest, Hungary, paper no. 108. Aug
15. Li H, Chen Z (Jun. 2008) Overview of generator topologies for wind turbines. *IET Proc Renew Power Gener* 2(2):123–138
16. Mihet-Popa L, Blaabrier F (2004) Wind turbine generator modeling and simulation where rotational speed is the controlled variable. *IEEE Trans Ind Appl* 40(1):3–10
17. Chowary BH, Chellapilla S (2006) Doubly-fed induction generator for variable speed wind power generation. *IEEE Trans Electric Power Sys Res* 76:786–800
18. Roldán-Pérez J, García-Cerrada A, Ochoa-Giménez M, Zamora-Macho JL (Oct. 2016) On the power flow limits and control in series-connected custom power devices. *IEEE Trans Power Electron* 31(10):7328–7338
19. Eklashossain, Padmanaban S (2018) Analysis and mitigation of power quality issues in distributed generation systems using custom power devices. *IEEE Access* 6:16816–16833
20. Kumar PKA, Vivekanandan S, Kumar CK, and Chinnaiyan VK (2016) Neural network tuned fuzzy logic power system stabilizer design for SMIB. 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, pp 446–451. <https://doi.org/10.1109/IC3I.2016.7918006>
21. Kumar PKA, Uthirasamy R, Saravanan G, Ibrahim AM (2016) AGC performance enhancement using ANN. 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, pp. 452–456. <https://doi.org/10.1109/IC3I.2016.7918007>

Chapter 3

Detection and Classification of Power Quality Events Using Wavelet Energy Difference and Support Vector Machine



Arun Kumar Puliyadi Kubendran and L. Ashok Kumar

Abstract Detection and classification of power quality events (PQE) to improve the quality of electric power is an important issue in utilities and industrial factories. In this paper, an approach to classify PQE with noise based on wavelet energy difference and support vector machine (SVM) is presented. Here PQE signals are decomposed into ten layers by db4 wavelet with multi-resolution. Energy differences (ED) of every level between PQE signal and standard signal are extracted as eigenvectors. Principal component analysis (PCA) is adopted to reduce the dimensions of eigenvectors and find out the main structure of the matrix, which forms new feature vectors. Then these new feature vectors are divided into two groups, namely training set and testing set. The method of cross-validation is used for the training set to select the optimal parameters adaptively and construct the training model. Also, the testing set is substituted into the training model for testing. Finally, the proposed method results are compared with S-transform (ST)- and Hilbert-Huang transform (HHT)-based PQE classification to verify the accuracy of classification. The results demonstrated show that the proposed method has high accuracy, strong resistance to noise, and fast classification speed and is suitable for the detection and classification of PQE.

Keywords Power quality · Multi-resolution analysis · Pattern recognition · Wavelet energy difference · Principal component analysis · S-transform · Hilbert-Huang transform · Support vector machine

A. K. Puliyadi Kubendran (✉)
Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India

L. Ashok Kumar
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

Abbreviations

ED	energy differences
HHT	Hilbert-Huang transform
MRA	multi-resolution analysis
PCA	principal component analysis
PQE	power quality events
ST	S-transform
SVM	support vector machine

3.1 Introduction

The increase in electric power demand in india can be meet out using the renewable energy system, that is connected in parallel with the existing traditional power grid. However, the increasingly large power system as well as a variety of power electronic devices, the impact and the non-linear load cause a series of power quality problems such as voltage swells, voltage sag, voltage interruption, harmonics, transient pulse, transient oscillation and voltage flicker. Power quality monitoring and analysis is the prerequisite for finding and controlling the power quality problems. There are Fourier transform, the dq-transform, wavelet transform, etc. about feature extraction in power quality events (PQE) recognition system [1–7]. In view of the wavelet transform of the time-frequency localization characteristics, the wavelet energy difference is extracted as the feature vectors in the paper. Artificial neural networks, fuzzy logic, expert systems, support vector machine (SVM), etc. are applied in the classification of PQE [8–13]. SVM is introduced to apply to classification of PQE in paper. SVM improves the generalization ability of learning methods. It proposes a method to identify PQE with noise based on PCA for reducing the dimensions of eigenvectors and SVM of optimizing the nuclear parameters. Firstly, it gets waveforms and samples according to the mathematical models of PQE, extracts the wavelet energy difference (ED) coefficients using the multi-resolution wavelet transform and then reduces the dimensions of the wavelet coefficients using PCA which removes irrelevant data and makes the remaining data as the feature vectors. Finally, the feature vectors are inputted to SVM classifier to identify PQE.

The article is organized as follows in Sect. 3.2; overview of principal component analysis is explained. A basic of SVM and HHT is described in Sects. 3.3 and 3.4 and the recommended SVM classifiers are investigated by means of various noise conditions and the outcomes are presented in Sect. 3.5 and followed by the conclusion.

3.2 Principal Component Analysis

The principal component analysis (PCA) is a statistic method which holds the inconsistency of things. PCA can parse out the main factors from multi-things, reveal the nature of things and simplify the complex issues. Projecting the high-dimensional data onto a lower dimensional space is the main purpose of PCA. PCA is mainly used for reducing the dimension of data.

3.2.1 Steps to Reduce the Dimension of the Matrix with PCA

1. To calculate covariance matrix $S_{n \times n}$ of the sample matrix $X_{m \times n}$.
2. To calculate the eigenvectors of covariance matrix $S_{n \times n}$ and the eigenvalues of the eigenvectors. The eigenvalues are arranged from the larger to the smaller. The eigenvectors are e_1, e_2, \dots, e_N , and the eigenvalues are x_1, x_2, \dots, x_N , for $i = 1, 2, \dots, N$.
3. To calculate the contribution rate and the cumulative contribution rate of the covariance matrix, then the formulas are as follows:

$$\theta_i = x_i / \sum_{n=1}^N x_n, i = 1, 2, \dots, N \quad (3.1)$$

$$\Theta_r = \sum_{i=1}^r \theta_i, r = 1, 2, \dots, N \quad (3.2)$$

where θ_i is the contribution rate of the i th column vector of the covariance matrix $S_{n \times n}$, when Θ_r is the cumulative contribution rate of the first r column vectors.

4. To determine the dimension of the projection matrix (or it is called PCA matrix), and the dimension is r if $\Theta_r > 70\%$, $r < n$.
5. To take the first r -dimension of the covariance matrix as the projection matrix $S_{n \times r}$. $T_{n \times n}$ is the covariance matrix of the needed to down-dimensional matrix $Y_{s \times n}$, and the result of $T_{n \times n}$ multiplied by $S_{n \times r}$ is the final matrix we need. Namely, $T_{n \times n} \times S_{n \times r}$.

3.3 Support Vector Machine

A support vector machine is a new learning machine whose generalization ability of learning machines (both neural network and SVM) depends on the structural risk minimization (SRM) inductive principle. For controlling the generalization ability of

learning machines, an inductive principle for minimizing the risk functional using a small sample of training instances is preferred. Vapnik found a new pattern recognition method, namely, the support vector machine (SVM), which shows a great advantage to solve the problem of the small sample, non-linear and high-dimensional pattern recognition. Looking for the final discriminate function of the optimal hyper-plane is the ultimate goal of SVM. The optimal hyper-plane, found during a training phase, makes the smallest number of training errors.

3.4 Hilbert-Huang Transform

The HHT investigation technique is a better tool for nonuniform and unstationary time series electric power signal processing. It is mainly based on EMD and the Hilbert spectrum inquiry. First, EMD is applied to decompose the nonuniform, unstationary time series electric power signals into dissimilar frequency modules, starting from maximum frequency to minimum frequency. These modules are termed as intrinsic mode functions (IMF).

The EMD decomposes each complex signal into different, simple and non-sinusoidal signals; also, an IMF is a function that fulfils two circumstances: First, in the entire data set, the integer of extreme and zeros must either equal or differ at most by one. Second, the mean values of the envelopes defined by the local maximum and minimum are relative time symmetric and only one single wave mode exists between two IMF zeros in each wave cycle. The significance of the first circumstance is obvious, which is related to the narrow band condition of the fixed Gaussian process. The intuitive meaning is that no minimum value greater than zero and no maximum value less than zero can be found in the IMF. The second condition is the result of modifying a traditional global condition to a local condition, which is essential to confirm that unwanted fluctuations due to asymmetry of waveforms do not occur in the instantaneous frequency. Preferably, the condition should be that the data have a local mean of zero. For unstationary signals, the calculation of the local mean is related to the local time scale, but this cannot be determined. Therefore, the local mean of the envelope defined by the maximum and minimum values is used in the EMD instead of the actual mean to ensure the local symmetry of each intrinsic mode function. Although this is an approximation, it avoids determining the local average time scale. Although this approach introduces some aliasing due to the non-linear deformation of the signal, the resulting non-linear effect is much less than the non-stationary one.

HHT is reported to detect and investigate the multi-complex electric PQD signal in any power system and eradicates more limitations present in the other transforms when compared to HHT [14–17].

3.5 Power Quality Analysis Using SVM

3.5.1 Mathematical Models of PQE

According to the IEEE for the definition of various types of PQE, eight models of power quality are established [18–20] in Table 3.1. T is frequency cycle, and the sampling frequency is 6400 HZ, when t_1 is 0.071 s, t_2 is 0.142 s. Gaussian random white noise with SNR = 25 db is added to verify the robustness of SVM algorithm. The PQED waveforms are shown in Fig. 3.1

Table 3.1 Mathematical models of PQE

PQE	Expression
Normal	$u(t) = \sin \omega 0t$
Swells	$u(t) = \{1 + \alpha[u(t - t_1) - u(t - t_2)]\} \sin \omega 0t$
Sag	$u(t) = \{1 + \alpha[u(t - t_1) - u(t - t_2)]\} \sin \omega 0t$
Interruption	$u(t) = \{1 + \alpha[u(t - t_1) - u(t - t_2)]\} \sin \omega 0t$
Harmonics	$u(t) = \{1 + \alpha[u(t - t_1) - u(t - t_2)]\} \sin \omega 0t$
Pulse	$u(t) = \sin \omega 0t + \alpha 3 \sin \omega 0t + \alpha 5 \sin \omega 0t + \alpha 7 \sin \omega 0t$
Oscillation	$u(t) = \sin \omega 0t + \alpha e^{-c(t - t_1)}[u(t - t_1) - u(t - t_2)] \sin \beta \omega 0t$
Flicker	$u(t) = \left[1 + m \sum_i^3 \frac{1}{i} \sin (2\pi if 1t) \right] \sin \omega 0t$

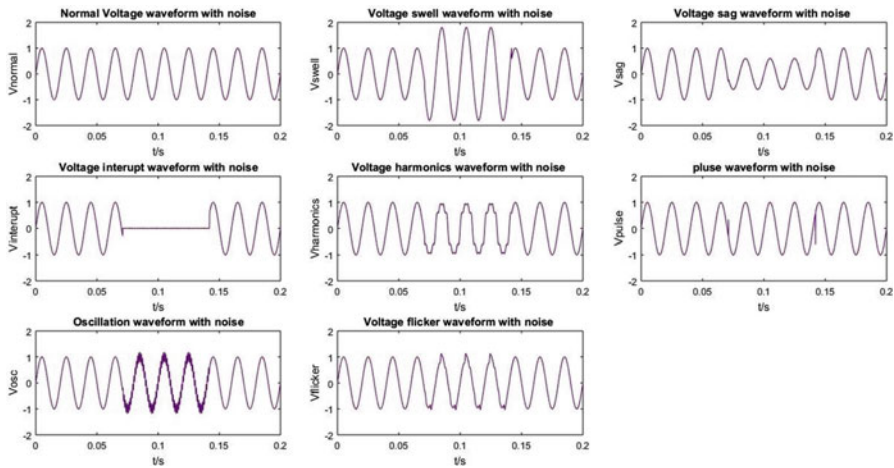


Fig. 3.1 Different power quality events

3.5.2 Multi-resolution Analysis (MRA)

The multi-resolution analysis (MRA) is defined that it decomposes the signal into a high-frequency X component and a low-frequency X component which is decomposed into a sub-high-frequency X component and a low frequency X component by the high-pass filters and low-pass filters. Subsequent decomposition of low frequency component yields high-frequency component coefficients of the different scales and the low-frequency component coefficient of the lowest scale. The characteristics of the original waveform can be reflected in different scales after the MRA decomposition. Based on this observation, we can construct the feature vector to detect different kinds of PQ events.

3.5.3 Energy Difference of MRA of PQE

Taking into account the voltage swells, transient pulse and flicker, db4 wavelet whose length of filter coefficient 4, is used for 10-layer decomposition of PQE. As shown in Fig. 3.2, the changes of all waveforms appear between the fifth and the eighth layer, and the seventh layer of the waveforms has the greatest change. The ED waveforms of transient pulse and voltage flicker are raised, and others are sag. The ED between the different PQE types in each band are also not the same, which can be extracted as the feature vectors to recognize and classify PQE.

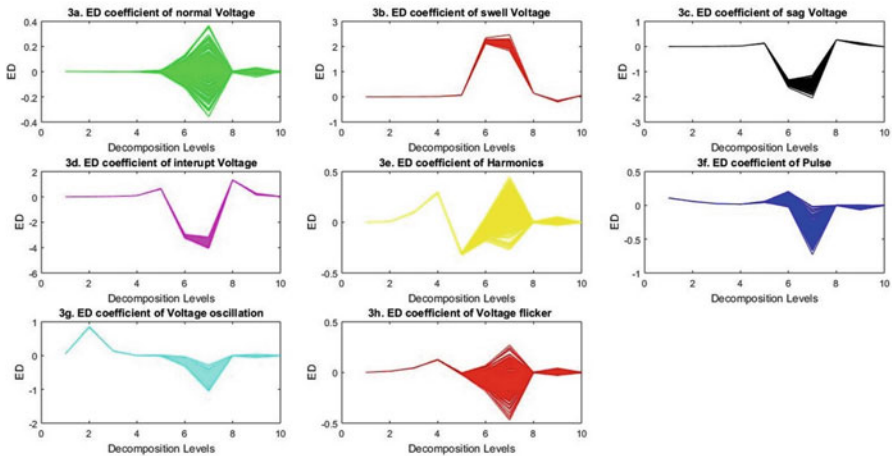


Fig. 3.2 Energy difference coefficient of various power quality events

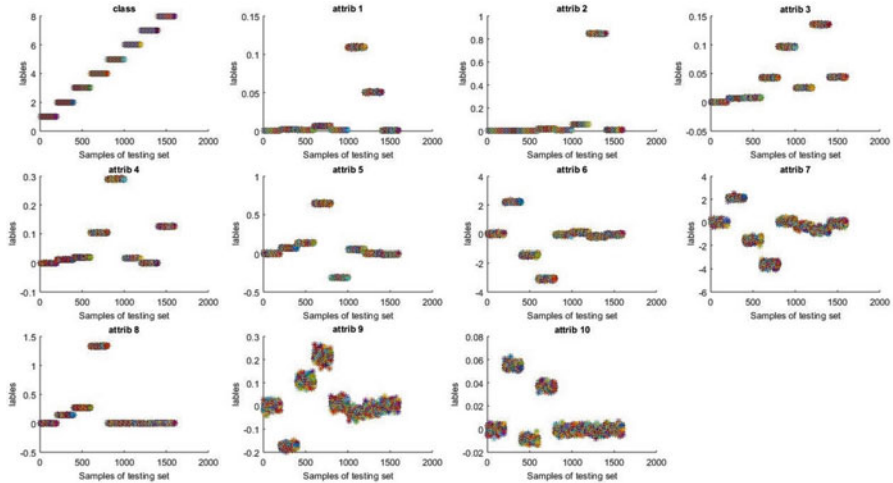


Fig. 3.3 Accurate diagram of actual and predictable classification for testing

3.5.4 Optimization of Parameters and Classification of Testing Set

The classification result of testing set shown in Fig. 3.3 is the classification result of the original data before reducing dimension and (b)–(h) are the classification results from 2 to 10 after reducing dimension. The specific classification of eight kinds of PQE can be shown intuitively in every figure. By comparing the predictable classification to the actual classification, it can clearly show that which samples were correctly classified, which samples were incorrectly classified. According to the statistical quantity of samples correctly and incorrectly classified, the classification accuracy of every disturbance can be calculated [21–25]. The probability of misclassification increases and the classification accuracy rate reduces as the reduction of the dimension of testing set by comparing these figures.

3.6 Conclusion

The proposed method is a simple and effective methodology for detection and classification of power quality disturbances. SVM interprets individual disturbance with seven distinguished time-frequency statistical features. Using a rule-based decision tree, seven power disturbances are classified and the relative energies between them have been observed. By experimental work on power quality problems such as sag, swell, interruption, harmonics, pulse, oscillatory transients and flicker disturbances are classified using SVM. Similarly, the classification is done in

Table 3.2 Comparison result of PQE classifier

Number of test samples	Types of PQ event	SVM classifier	S-transform classifier	HHT classifier
200	Sag	199	195	197
200	Swell	195	193	196
200	Interruption	199	194	195
200	Harmonics	195	192	194
200	Pulse	194	186	190
200	Flicker	190	189	189
200	Oscillatory Transients	198	194	196
<i>Overall success rate</i>		97.857%	95.929%	96.929%

S-transform and HHT using the rule-based decision tree provided in Table 3.2, and it is observed that the results obtained are better in SVM-based classifier when compared to S-transform [26, 27] and HHT-based classifier.

Here several power quality problems have been analysed; also the SVM, S-transform and HHT classifier provide an interesting and significant tool in detecting and classifying the problem. SVM has proven abilities to develop fully automated monitoring systems equipped with online classifier.

References

1. Santoso S, Powers EI, Grady WM, Parsons AC (2000) Power quality disturbance waveform recognition using wavelet-based neural classifier—part 1: theoretical foundation. *IEEE Trans Power Deliv* 15(1):222–228
2. Gaing ZL (2004) Wavelet-based neural network for power disturbance recognition and classification. *IEEE Trans Power Deliv* 19(4):1560–1568
3. He H, Starzyk JA (2006) A self-organizing learning array system for power quality classification based on wavelet transform. *IEEE Trans Power Deliv* 21(1):286–295
4. Kopsinis Y, McLaughlin S (Apr. 2009) Development of EMD based denoising methods inspired by wavelet thresholding. *IEEE Trans Signal Process* 57(4):1351–1362
5. Masoum MAS, Jamali S, Ghaarzadeh N (2010) Detection and classification of power quality disturbances using discrete wavelet transform and wevelet networks. *IET Sci Meas Technol* 4(4):193–205
6. Tse NCF, Chan JYC, Wing-Hong L, Lai LL (Dec. 2012) Hybrid wavelet and Hilbert transform with frequency-shifting decomposition for power quality analysis. *IEEE Trans Instrum Meas* 61(12):3225–3233
7. Eristi B, Yildirim Ö, Erişti H, Demir Y (2018) A new embedded power quality event classification system based on the wavelet transform. *Int Trans Electr Energy Syst.* <https://doi.org/10.1002/etep.2597>
8. Chikuri MV, Dash PK (2004) Multiresolution S-transform-based fuzzy recognition system for power quality events. *IEEE Trans Power Deliv* 19(1):323–330
9. Mishra S, Bhende CN, Panigrahi BK (2007a) Detection and classification of power quality disturbances using S-transform and probabilistic neural network. *IEEE Trans Power Deliv* 23(1):280–287

10. Bhende CN, Mishra S, Panigrahi BK (2007b) Detection and classification of power quality disturbances using s-transform and modular neural network. *Electr Power Syst Res* 78(1):122–128
11. Uyar M, Yildirim S, Gencoglu MT (2008) An expert system based on S-transform and neural network for automatic classification of power quality disturbances. *Expert Syst Appl* 36(3):5962–5975
12. Hooshmand R, Enshae A (2010) Detection and classification of single and combined power quality disturbances using fuzzy systems oriented by particle swarm optimization algorithm. *Electr Power Syst Res* 80:1552–1561
13. Decanini JGMS, Tonelli-Neto MS, Malange FCV, Minussi CR (2011) Detection and classification of voltage disturbances using a Fuzzy-ARTMAP-wavelet network. *Electr Power Syst Res* 81:2057–2065
14. Shukla S, Mishra S, Singh B (Oct. 2009) Empirical mode decomposition with Hilbert transform for power quality assessment. *IEEE Trans Power Deliv* 24(4):2159–2165
15. Afroni MJ, Sutanto D, Stirling D (Oct. 2013) Analysis of non-stationary power quality waveforms using iterative Hilbert huang transform and SAX algorithms. *IEEE Trans Power Deliv* 28(4):2134–2144
16. Chen X, Li K, Xiao J (2018) Classification of power quality disturbances using dual strong tracking filters and rule-based extreme learning machine. *Int Trans Electr Energy Syst*:e2560. <https://doi.org/10.1002/etep.2560>
17. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and hilbert spectrum for nonlinear and non stationary time series analysis. *Proc R Soc London Ser A* 454:903–995
18. Samantaray SR, Dash PK, Panda G (2006) Power system events classification using pattern recognition approach. *Int J Emerg Electr Power Syst* 6(1):1–16
19. Dash PK, Nayak M, Lee IWC (2007) Time series pattern recognition using S-transform and soft computing. *Int J Knowl-Based Intell Eng Syst* 11(6):355–370
20. Panigrahi BK, Dash PK, Reddy JBV (2008) Hybrid signal processing and machine intelligence techniques for detection, quantification and classification of power quality disturbances. *Eng Appl Artif Intell* 22(3):442–454
21. Barbosa BHG, Ferreira DD (2013) Classification of multiple and single power quality disturbances using a decision tree-based approach. *J Contr Autom Electr Syst* 24(5):638–648
22. Ferreira DD, de Seixas JM, Cerqueira AS (2015) A method based on independent component analysis for single and multiple power quality disturbance classification. *Electr Power Syst Res* 119:425–431
23. Kubendran AKP, Loganathan AK (2017) Detection and classification of complex power quality disturbances using S-transform amplitude matrix-based decision tree for different noise levels. *Int Trans Electr Energy Syst* 27(4):1–12. <https://doi.org/10.1002/etep.2286>
24. Saini M, Beniwal R (2018) Detection and classification of power quality disturbances in wind-grid integrated system using fast time-time transform and small residual-extreme learning machine. *Int Trans Electr Energy Syst* 28:e2519. <https://doi.org/10.1002/etep.2519>
25. Kumar PKA, Vijayalakshmi VJ, Karpagam J, Hemapriya CK (2016) Classification of power quality events using support vector machine and S-Transform. 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, pp. 279–284. <https://doi.org/10.1109/IC3I.2016.7917975>
26. Stockwell RG, Mansinha L, Lowe RP (1996) Localization of the complex spectrum: the S transform. *IEEE Trans Signal Process* 44(4):998–1001
27. Solar PV and wind energy conversion systems: an introduction to theory, modeling with MATLAB/SIMULINK, and the role of soft computing techniques, S Sumathi, LA Kumar, P Surekha, Springer, Green Energy and Technology, <https://doi.org/10.1007/978-3-319-14941-7>, ISBN: 9783319149400, 9783319149417 (online)

Chapter 4

PMSM Drive Using Predictive Current Control Technique for HVAC Applications



L. Ashok Kumar and V. Indragandhi

Abstract This work deals with Predictive Current Control (PCC) strategy using Field Oriented Control (FOC) for Permanent Magnet Synchronous Machines (PMSM), to improve the performance of the current loop. Based on the research conducted, FOC and PCC have good performance in PMSM drives. Hence a new method based on combined FOC-PCC is adopted in this project work. The main objectives of the FOC-PCC controller are to improve the dynamic characteristics of the drive and to reduce the current ripple. The simulation has been implemented with MATLAB/Simulink for a 750 W PMSM with rated speed of 3000 rpm and rated current of 4.8 A. The sampling time has been set to 20 μ s, and the DC link voltage is maintained at 180 V. The reference speed has been set to 3000 rpm and load torque is set to 2 Nm. The stator currents have been limited below its maximum allowable set limit during the dynamic operations such as starting and braking and also during speed reversals with the help of control technique.

Keywords Field oriented controller · Predictive current control · PMSM · Speed reversal · THD

Abbreviations

DTC	Direct Torque Control
FOC	Field Oriented Control
MPC	Model Predictive Control

L. Ashok Kumar (✉)
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

V. Indragandhi
Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: indragandhi.v@vit.ac.in

PCC	Predictive Current Control
PMSM	Permanent Magnet Synchronous Machines
THD	Total Harmonic Distortion

4.1 Introduction

Late years, different strategies for non-linear control hypothesis have been utilized as a part of the PMSM control framework, for example, adaptive control [1, 2], active disturbance rejection control [3], fuzzy control [4], predictive control [5], robust control [6, 7], variable structure control [8, 9] and so on. These progressed non-linear control strategies have enhanced the execution of the PMSM framework from various perspectives. Nonsingular Terminal Sliding Mode Control (NTSMC) not just has points of interest in quick powerful reaction, limited time meeting and high control exactness; additionally it can dispense with the paranormal wonder of control system [10, 11].

Additionally sensorless control in light of the unsettling influence spectator has been examined in [12–15]. Moreover, a lot of research has been committed to enhance stack torque power for PMSM with mechanical sensors [16–20] and additionally their sensorless counterparts [21]. Be that as it may, the impact of the speed estimation blunder on torque unsettling influence dismissal and precise execution in sensorless strategies has to a great extent been overlooked.

In some of these techniques, new exchanging tables are acquainted with lessening the torque and fluxing swells in PMSM drives [22, 23]. In [24–27], prescient control strategies are used to decrease torque ripples by presenting a general investigative answer for estimation of the correct voltage vector; in this manner, the torque mistakes are wiped out towards the finish of each control period [28–30]. With a specific end goal to satisfy high-data transfer capacity prerequisites, the need of the use of the dead-beat controllers rather than the customary PI controllers is accounted for in [31]. Other than dead-beat controllers, some control techniques, for example, hysteresis, MPC, DTC [32] and their blends [33], have been executed to give quick present or torque reaction.

A discrete time robust predictive controller which is anything but difficult to execute and to play out the total strength investigation for a non-linear PMSM drive with instabilities is outlined in [33]. The execution of a PWM–MPC controller algorithm is offered for three-phase power systems, using the vector modulation preparation as a system to compute the best actuation from finite set current error expectations.

The main requirement in industries today is on the development of controller for drives which guarantees the best response. The properties of this controller must be better than other popular controllers such as Field Oriented Controller (FOC) and DTC. The inverter must provide smallest possible error in the controlled variables, minimization of switching losses, THD restrictions, good performance and improvement in overall efficiency of the drive. Taking these requirements into account, this work focuses on developing a control technique to meet the necessities.

4.2 Proposed Predictive Current Control

The system shown in Fig. 4.1 is implemented in MATLAB/Simulink for the analysis. This system is to be verified for better dynamic performance and faster control, so that this drive can be utilized for HVAC applications. At the point when executed in the controller, the calculation ought to consider the three assignments. In the first place, it is to foresee the conduct of the controlled variables for all exchanging states. Second, it ought to assess the cost capacity of every expectation. Third, it ought to choose the exchanging state that limits the cost function.

Execution of predictive models and control system may experience diverse troubles relying upon the kind of stage utilized. At the point when executed utilizing a fixed-point processor, unique consideration must be paid for the programming so as to get the best precision in the fixed-point portrayal of the variables. Then again, when executed utilizing a floating point processor, nearly a similar programming utilized for simulations can be utilized as a part of the research facility.

Contingent upon the unpredictability of the controlled framework, the quantity of computations can be noteworthy and will restrict the base sampling time. In the easiest case, predict current control, the count time is little, yet in different plans, for example, torque and flux control, the estimation time is the parameter which decides the permitted sampling time. The quantity of computations is specifically identified with the quantity of conceivable switching states. In case of three-phase system, two-level inverter to compute predictions for the eight-probable-switching state is not problematic, but in case of multilevel and multiphase systems, a dissimilar optimization technique must be considered to decrease the number of controls.

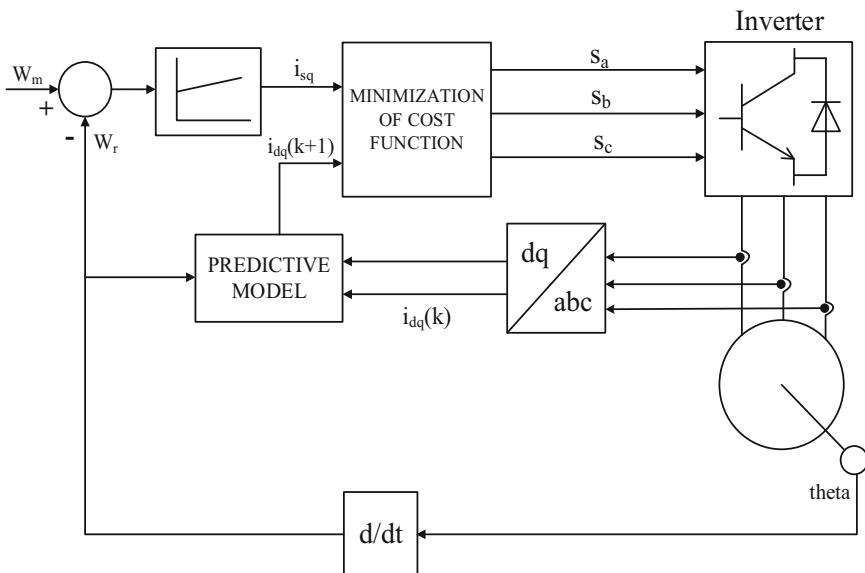


Fig. 4.1 Block diagram of FOC-PCC for PMSM

4.3 Results and Discussions

The simulation has been implemented with MATLAB 2013a/Simulink for a 750 W PMSM with a rated speed of 3000 rpm and a rated current of 4.8 A. The sampling time has been set to 20 μ s, and the DC link voltage is maintained at 180 V.

The reference speed is given as input to the comparator, where it is compared with the actual speed of the rotor. This error signal is given to a PI controller which converts the error to corresponding i_q^* , q axis reference current. The rotor position is estimated and this position θ_e is used to align the rotor flux with d axis. The i_q^* current reference is given as input to the predictive controller along with the measured values of stator currents $i_d(k)$, $i_q(k)$. The predictive controller will predict the future value of stator current for each of the voltage vector. The corresponding switching state for desired voltage vector is applied to the VSI. The DC link voltage to the inverter is maintained at 180 V. With rated load in the shaft, the speed reversal command is given at 0.25 s and the new set speed is -2000 rpm from 0.25 s. The three-phase voltages are given to the first block, which converts the three-phase voltage vectors to corresponding V_d and V_q components. From the model, the outputs obtained are i_d , i_q , rotor position (θ_e) and electromechanical torque (T_e). Figure 4.2 shows the rotor speed waveform for the PMSM drive.

From the speed waveform in Fig. 4.2, it is visible that PMSM takes 0.05 s to settle to 3000 rpm from rest. On the application of load at 0.1 s, the oscillation in the speed is also settled within 0.05 s. Similarly during speed reversal at 0.25 s, it can be inferred that the new speed is reached within 0.075 s. The speed ripple and the error

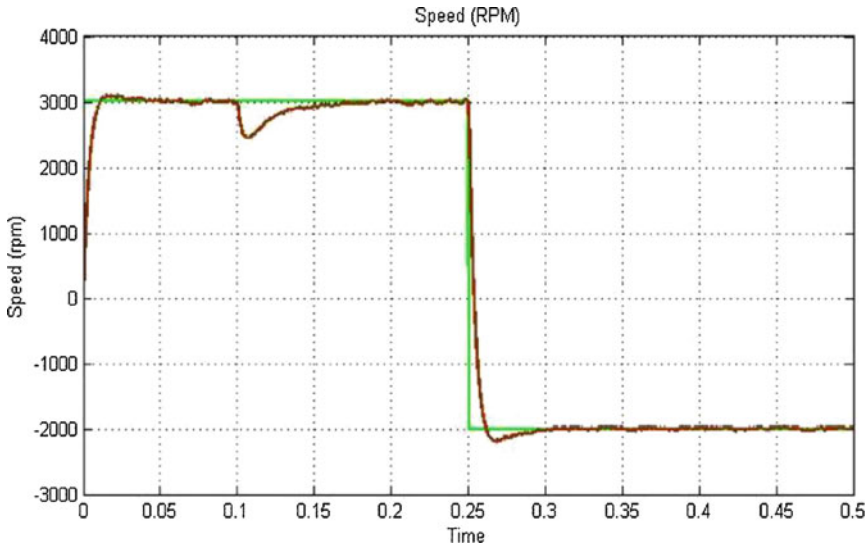


Fig. 4.2 Speed waveform of motor

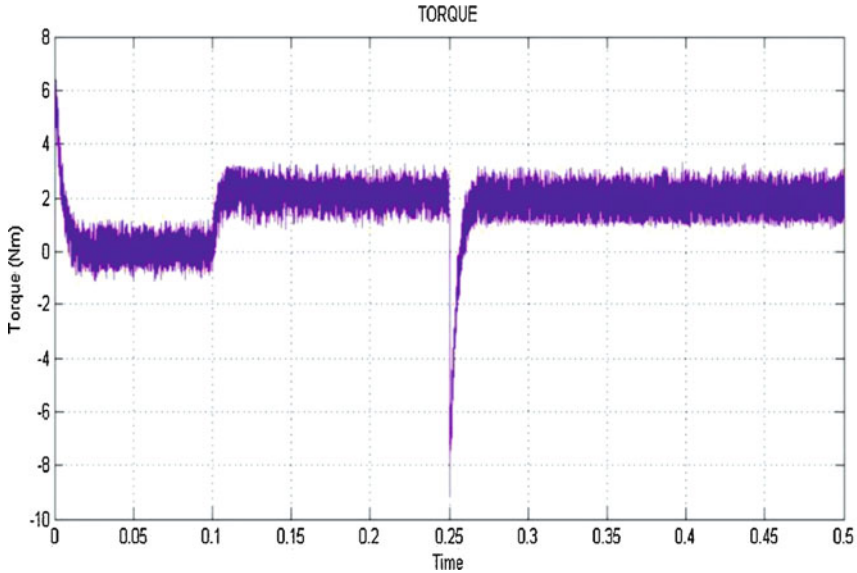


Fig. 4.3 Electromechanical torque of motor

between set speed and actual speed are also considerably reduced. Figure 4.3 shows the electromechanical torque variations of the PMSM. During starting, the motor torque is raised to about 3 Nm in order to achieve faster acceleration.

Figure 4.4 shows the d -axis and q -axis component of the stator currents. From the waveform it is clear that the id component is almost zero. This means $i_q = I_s$. Hence the maximum torque can be achieved with minimum stator current. This helps in the reduction in the stator copper loss. There is no need of id current as the rotor flux is created by permanent magnets in the rotor. The stator currents are shown in Fig. 4.5.

From Fig. 4.6, the current THD is restricted to 1.6%. During starting, oscillations and speed reversal, the stator currents are kept within the maximum set limit without considerably affecting the performance of the drive. Moreover, both the d and q components are decoupled. When the stator and rotor fluxes are 90 degrees apart, then the produced electromechanical torque in rotor is in maximum value. For this $id = 0$, and $i_q = I_s$. This is also achieved in the simulation, thereby optimizing torque by current ratio. The current ripple is limited with less THD in current waveform. Smooth speed control and fast response for changes in the input are also observed. Moreover, the stator current is kept within its maximum limit specified as 15 A.

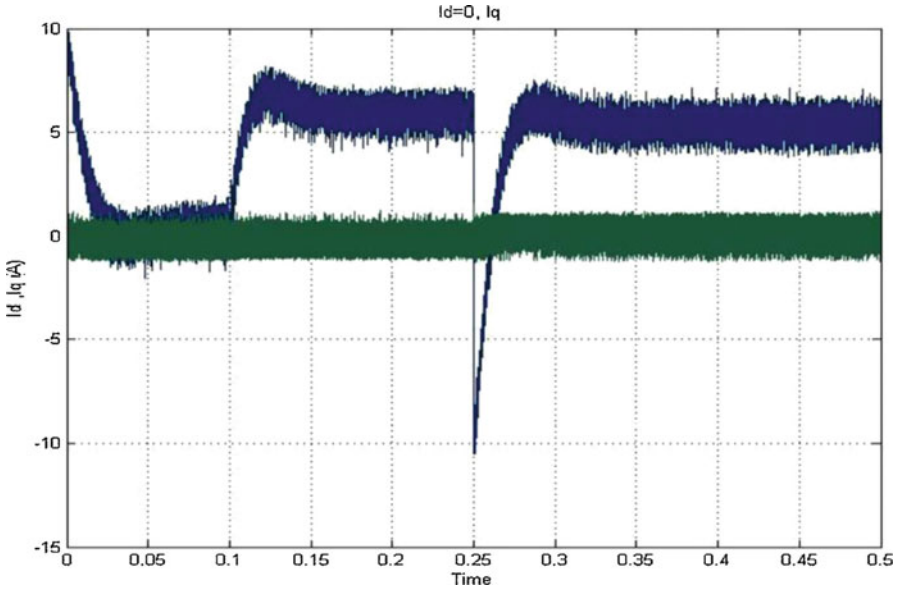


Fig. 4.4 D-axis and q-axis component of stator currents

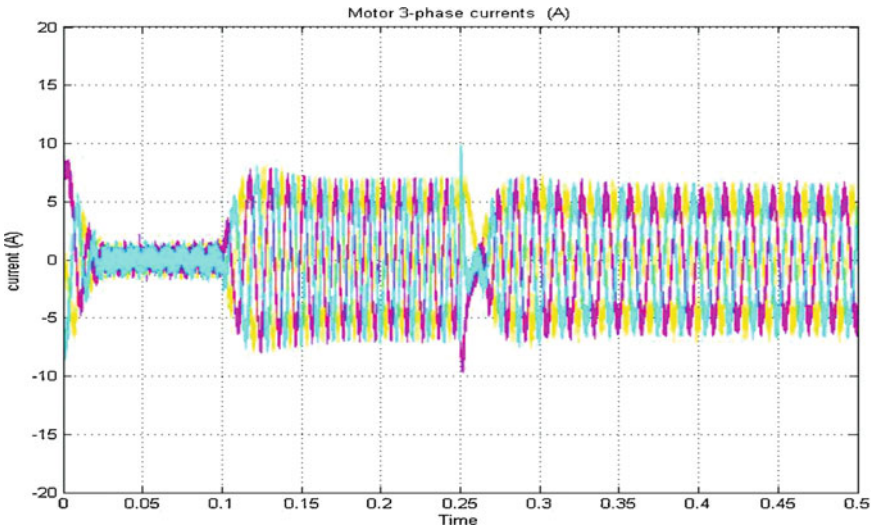


Fig. 4.5 Three-phase stator currents

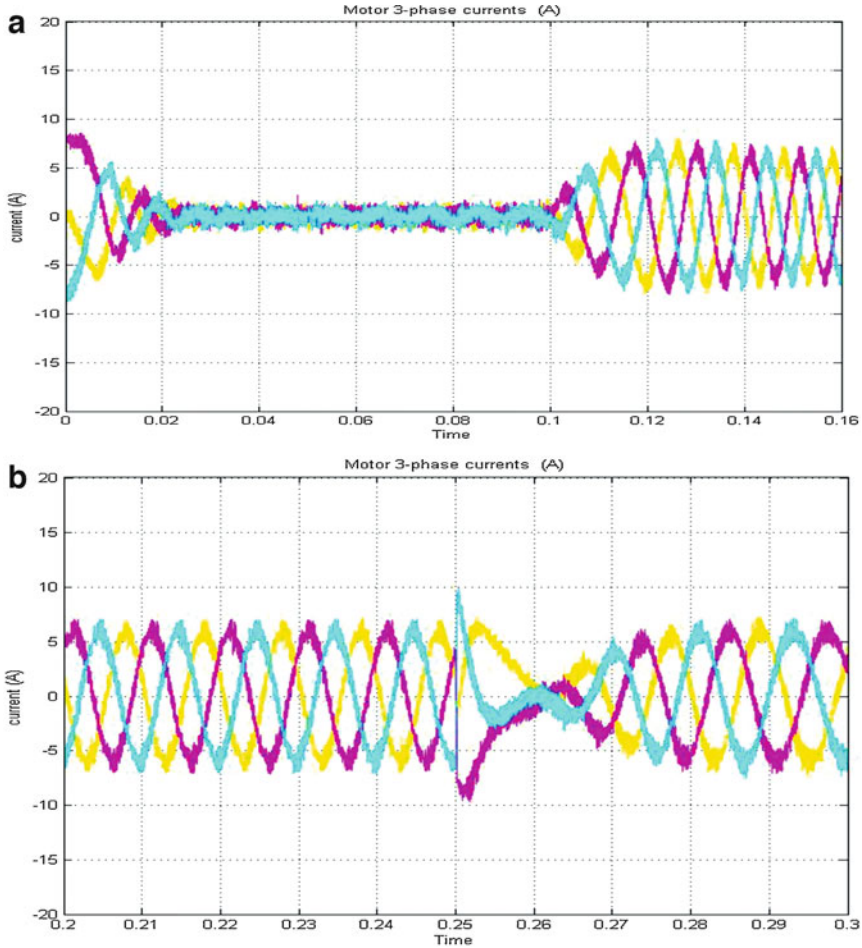


Fig. 4.6 Stator currents of motor during dynamic operations (a) starting and application of load (b) during speed reversal

4.4 Comparison of Predictive Current Control with Classical Control Techniques

The simulation of conventional FOC for PMSM of same specifications with same switching frequency of inverter is implemented in Simulink. The simulation has been carried for a total period of 0.5 s. A load torque of 2 Nm is applied to the motor at 0.1 s. The set speed is 3000 rpm up to 0.25 s. Speed reversal takes place at 0.25 s and the new speed is -2000 rpm (Fig. 4.7).

By comparing the results obtained from the simulation of conventional FOC with PCC-FOC, the following results can be concluded. The dynamic characteristic of the speed loop is better in case of FOC-PCC than conventional FOC, and the control in

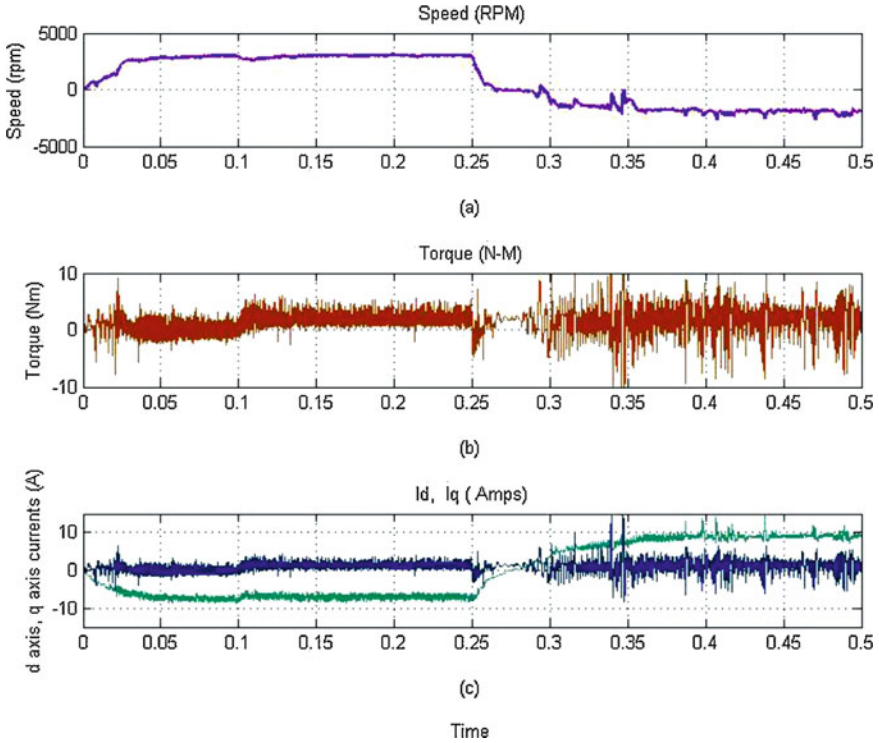


Fig. 4.7 Dynamic characteristics of motor for conventional FOC (a) speed, (b) torque and (c) I_d and I_q

speed is also smooth in FOC-PCC. Electromechanical torque ripple in the FOC-PCC is less than conventional FOC.

For the same load torque of 2 Nm, the stator current in FOC-PCC is 6 A, whereas in case of conventional FOC, it is 10 A. During the speed reversal, in case of FOC-PCC, the stator currents are maintained within the maximum set limit of 15 A, but in conventional FOC, there is no control over the maximum stator current. In FOC-PCC the I_d component of stator current is zero. But in case of conventional FOC, the magnitude of I_d is around 10 A. This will increase the copper losses in the stator. But the stator copper loss gets reduced in FOC-PCC technique. From the above inference, it is clear that in order to have a better drive for HVAC applications in the industries, which requires high efficiency, better control and faster response. The FOC-PCC is better when compared to conventional FOC for PMSM drive (Figs. 4.8a and 4.8b).

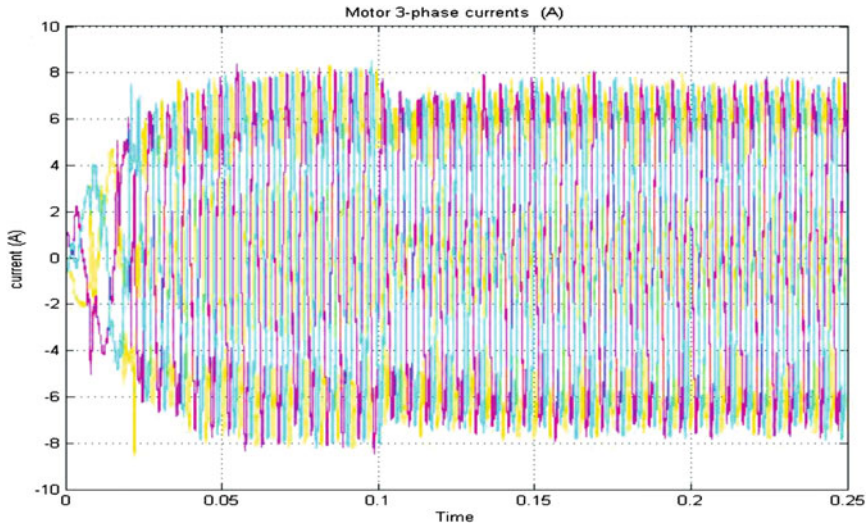


Fig. 4.8a Stator currents during starting of motor and application of load

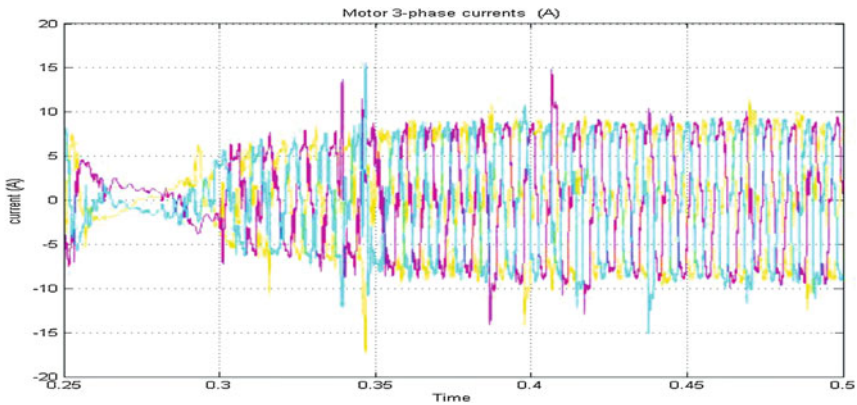


Fig. 4.8b Stator currents during the speed reversal

4.5 Conclusion

Another control procedure in view of ordinary FOC is recommended to give a straightforward and precise control strategy in PMSM drive. Likewise, the proposed control procedure utilizes predict current control to assess the required stator current precisely. A few numerical reproductions utilizing MATLAB have been done in

relentless state and transient-state operation conditions. As indicated by the outcomes, the proposed method can diminish torque swell and its transient states at mechanical torque varieties. The advantages of the proposed method are torque to current optimization, fast response and low harmonic current. Based on the results, this drive can have good performance characteristics for HVAC applications compared to traditional FOC drives.

References

1. Mu C, Xu W, Sun C (2014) Adaptive sliding mode control for the speed regulation of PMSMs with load change. In: Proceedings of International Conference on Electrical Machines and Systems, pp 549–555
2. Yi Y, Don Mahinda V, Rahman MA (2003) Implementation of an artificial-neural network-based real-time adaptive controller for an interior permanent magnet motor drive. *IEEE Trans Ind Appl* 39(1):96–104
3. Fu Y, Liu H, Pang Y (2009) Active disturbance rejection control for the airborne PMSM in direct drive EMA application. In: Proceedings of the International Conference Electronic Measurement and Instrument, pp 2346–2351
4. Li P, Yan X, Liu J (2010) Study on inverse system of permanent magnet synchronous motor by decoupled fuzzy sliding mode control. In: Proceedings of the International Conference on Intelligent Control Information Processing, pp 538–542
5. Preindl M, Schaltz E (2011) Sensorless model predictive direct current control using novel second-order PLL observer for PMSM drive systems. *IEEE Trans Ind Electron* 58 (9):4087–4095
6. Lei G, Wang T, Zhu JG, Guo YG, Wang SH (2015) System-level design optimization methods for electrical drive systems: robust approach. *IEEE Trans Ind Electron* 62(8):4702–4713
7. Lei G, Guo YG, Zhu JG, Wang TS, Chen XM, Shao KR (2012) System level six sigma robust optimization of a drive system with PM transverse flux machine. *IEEE Trans Magn* 48 (2):923–926
8. Bernardes T, Montagner V, Grundling H (2014) Discrete-time sliding mode observer for sensorless vector control of permanent magnet synchronous machine. *IEEE Trans Ind Electron* 61(4):1679–1691
9. Zhang B, Pi Y (2011) Enhanced sliding-mode control for permanent magnet synchronous motor servo drive. *Proc Int Conf Contr Decis Conf*:122–126
10. Xu W, Jiang Y, Mu C (2015) Non-singular terminal sliding mode control for speed regulation of permanent magnet synchronous motor with parameter uncertainties and torque change. *Proc Int Conf Electr Mach Syst*:2034–2039
11. Mu C, Sun C, Xu W (2016) On switching manifold design for terminal sliding mode control. *J Frankl Inst* 353(7):1553–1572
12. Lee K-W, Ha J-I (2012) Evaluation of back-EMF estimators for sensorless control of permanent magnet synchronous motors. *J Power Electron* 12(4):605–614
13. Kim JS, Sul SK (1995) High performance PMSM drives without rotational position sensors using reduced order observer. Thirtieth IAS Annual Meeting, IAS '95, Conference Record of the 1995 IEEE industry applications conference, 1995, 1, pp 75–82
14. Tomita M, Senjyu T, Doki S et al (1998) New sensorless control for brushless DC motors using disturbance observers and adaptive velocity estimations. *IEEE Trans Ind Electron* 45 (2):274–282

15. Morimoto S, Kawamoto Sanada KM, Takeda Y (2002) Sensorless control strategy for salient-pole PMSM based on extended EMF in rotating reference frame. *IEEE Trans Ind Appl* 38 (4):1054–1061
16. Chen Z, Tomita M, Doki S et al (2003) An extended electromotive force model for sensorless control of interior permanent magnet synchronous motors. *IEEE Trans Ind Electron* 50 (2):288–295
17. Solsona J, Valla MI, Muravchik C (2000) Nonlinear control of a permanent magnet synchronous motor with disturbance torque estimation. *IEEE Trans Energy Convers* 15(2):163–168
18. Sira-Ramirez H, Linares-Flores J, Garcia-Rodriguez C et al (2014) On the control of the permanent magnet synchronous motor: an active disturbance rejection control approach. *IEEE Trans Control Syst Technol* 22(5):2056–2063
19. Lee, K, Choy, I, Back J et al (2011) Disturbance observer based sensorless speed controller for PMSM with improved robustness against load torque variation. 2011 IEEE eighth International Conference on Power Electronics and ECCE Asia (ICPE ECCE), Jeju, Korea, May 2011, pp 2537–2543
20. Xia C, Wang S, Wang Z, Shi T (2016) Direct torque control for VSIPMSMs using four-dimensional switching-table. *IEEE Trans Power Electron* 31(8):5774–5785
21. Beerten J, Verdeccken J, Driesen J (2010) Predictive direct torque control for flux and torque ripple reduction. *IEEE Trans Ind Electron* 57(1):404–412
22. Zhu H, Xiao X, Li Y (2012) Torque ripple reduction of the torque predictive control scheme for permanent-magnet synchronous motors. *IEEE Trans Ind Electron* 59(2):871–877
23. Cho Y, Lee KB, Song JH, Lee YI (2015) Torque-ripple minimization and fast dynamic scheme for torque predictive control of permanent magnet synchronous motors. *IEEE Trans Power Electron* 30(4):2182–2190
24. Xie W et al (2015) Finite-control-set model predictive torque control with a deadbeat solution for PMSM drives. *IEEE Trans Ind Electron* 62(9):5402–5410
25. Vafaie MH, Dehkordi BM, Moallem P, Kiyomarsi A (2016) A new predictive direct torque control method for improving both steady-state and transient-state operations of the PMSM. *IEEE Trans Power Electron* 31(5):3738–3753
26. Zhang Y, Zhu J (2011) A novel duty cycle control strategy to reduce both torque and flux ripples for DTC of permanent magnet synchronous motor drives with switching frequency reduction. *IEEE Trans Power Electron* 26(10):3055–3067
27. Ren Y, Zhu ZQ, Liu J (2014) Direct torque control of permanent-magnet synchronous machine drives with a simple duty ratio regulator. *IEEE Trans Ind Electron* 61(10):5249–5258
28. Xia C, Zhao J, Yan Y, Shi T (2014) A novel direct torque control of matrix converter-fed PMSM drives using duty cycle control for torque ripple reduction. *IEEE Trans Ind Electron* 61 (6):2700–2713
29. Springob L, Holtz J (1998) High-bandwidth current control for torque ripple compensation in PM synchronous machines. *IEEE Trans Ind Electron* 45(5):713–721
30. Lee JS, Lorenz RD (2014) Deadbeat direct torque and flux control of IPMSM drives using a minimum time ramp trajectory method at voltage and current limits. *IEEE Trans Ind Appl* 50 (6):3795–3804
31. Geyer T (2013) Model predictive direct torque control: derivation and analysis of the state-feedback control law. *IEEE Trans Ind Appl* 49(5):2146–2157
32. Türker T, Buyukkeles U, Bakan AF (2016) A robust predictive current controller for PMSM drives. *IEEE Trans Ind Electron* 63(6):3906–3914
33. Belda K e, Vošmik D (2016) Explicit generalized predictive control of speed and position of PMSM drives. *IEEE Trans Ind Electron* 63(6):3889

Chapter 5

Grid-Connected 5 kW Mono-crystalline Solar PV System



L. Ashok Kumar, Sheeba Babu, and V. Indragandhi

Abstract Solar power industry in the country is growing rapidly. As of last month (September 2017), the country's aggregate solar capacity is 16.20 GW. The country increased its solar power production capability to about fourfold from 2650 MW in the month of May 2014 to 12,289 MW in the month of March 2017. The paper aims to offer tools along with strategy so as to ascertain that solar PV power systems are correctly specified and installed, giving an efficient arrangement which works to its rated capacity. The paper addresses solar PV systems set up on rooftop that are interconnected to grid. Statistical data studies reveal that about 10–20% of recently set-up solar PV systems have major drawbacks in installation which has led to performance deterioration, significantly reducing output. Proper sizing and orientation of the solar PV panel to obtain maximum electrical power and energy output are important. Poor public opinions resulting from improperly installed solar PV system will adversely affect the whole solar industry. In this paper, a 5 kW mono-crystalline solar PV system design analysis is carried out. Hardware results of 5 kW mono-crystalline solar PV installation are also presented.

Keywords Solar PV system · Grid-interactive inverter · Balance of system · MPPT technique · Power conditioning unit · Net metering

L. Ashok Kumar (✉)

Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

S. Babu

PSG College of Technology, Coimbatore, Tamil Nadu, India

V. Indragandhi

Vellore Institute of Technology, Vellore, Tamil Nadu, India
e-mail: indragandhi.v@vit.ac.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_5

Abbreviation

BoS	Balance of Systems
CEA	Technical Standards for Connectivity Distribution Generation Resources
DCDB	DC Distribution Box
DPB	Distribution Panel Board
JB	Junction Box
MOVs	Metal Oxide Varistors
PCU	Power conditioning unit
SPDs	Surge Protection Devices
STC	Standard Test Conditions

5.1 Introduction

The main reason for the rapid increase in solar PV system is owing to the following: good location for solar resource, climatic change, depletion of fossil fuel, energy security, etc. Moreover, the simplicity of installation and the ease and accessibility of the products and technology have made it increasingly appealing for consumers [1–3]. In a simple solar PV arrangement, price of the balance of systems (BoS) is considerable to module cost. The BoS consists of every unit in the solar PV arrangement excluding the panels like conductors, switchgears, frames, inverter, and surge protectors [4–7]. Basically with these above-mentioned units, only one can reduce expenses, modify solar PV arrangement and increase efficiency. So, proper combination of BoS elements is important for the good operation and the consistency of the solar photovoltaic arrangement. But as a matter of fact, it is mostly fully neglected and carelessly integrated [8]. This paper attempts to present the proper sizing methods and selection criteria for the solar PV system. A system having BoS which is properly put together can lead to added advantages like improved lifespan having minimum maintenance cost over its lifespan and improved efficiency.

The position and placement of PV system affect the operation and output of the system either favourably or adversely. The operations of such systems have to be analysed for different climatic regions so as to get a better picture of its nature of response at various regions [8–12]. PV array helps to convert energy in the sun's radiation to electrical energy, which is DC in nature. This DC power is then transformed to AC power by using an inverter. The AC power thus generated is given to load or is given to utility grid via main AC distribution board. According to analysis, the yield of solar PV unit is very much based on the PV module technology, its position and placement and installation methods which are followed. Not only these but also some ecological factors like radiation and atmospheric temperature are essential for designing the system capacity [13–15].

The power generated by a grid-tied PV system depends on (i) location, (ii) power converter parameters and (iii) the synchronization of grid, which relies on the energy generated by the power converter and on utility grid stability. One among the best

advantages of grid-connected PV system is that it does not need any storage device, which greatly reduces cost and size of system. But it comes with a trade-off due to its requirement of stable grid. If grid is not available, the system will not work even if the sun is shining bright. This paper presents the operational result of solar photovoltaic unit installed at PSG College of Technology, Coimbatore, for over a year.

5.2 Solar PV System Elements

A basic grid-interactive solar photovoltaic arrangement composes of solar PV module along with BoS (Fig. 5.1). The devices, elements and equipment that changes DC power generated by solar modules to AC power are generally referred to as BoS. Very often, all elements of PV system except the solar array come under BoS. Along with the power conditioning unit (PCU) and its accessories, this includes conductors, switchgears, protective devices, monitoring devices, frames and supporting structures [3].

BoS elements consist of major share of devices that roughly make up 10–50% of solar system component procurement expense along with expenses for setting up the

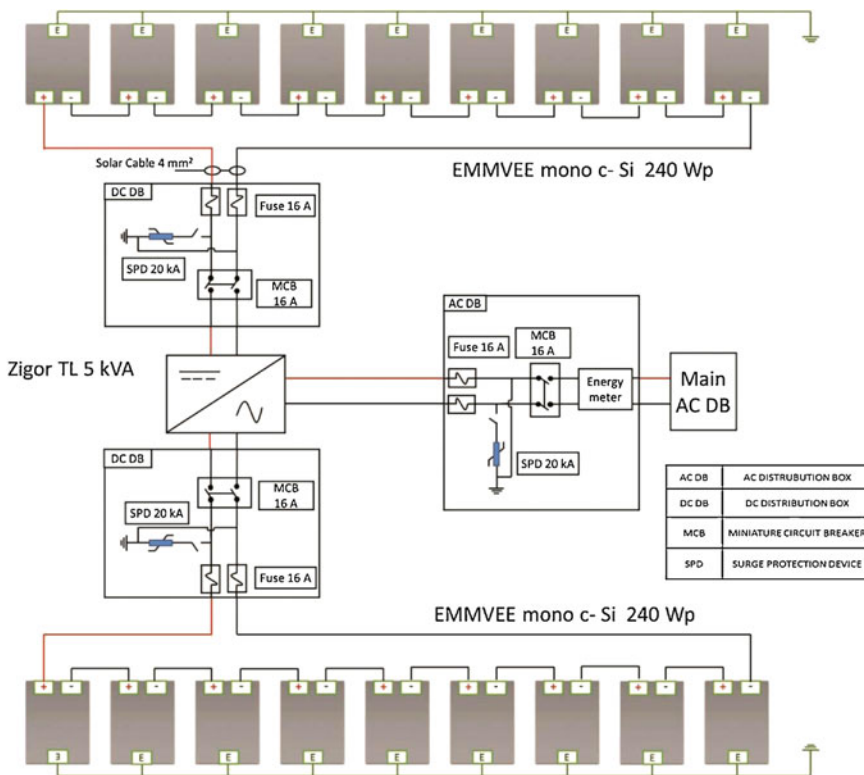


Fig. 5.1 Schematic arrangement of grid-interactive solar system at the site

system; it as well accounts for the lion share of repair needs. Basically it is through the BoS elements that one can manage expense, improve efficiency and modernize PV systems. With improvement in technology, prices are gradually decreasing with respect to solar modules and inverters [9].

5.2.1 Solar PV Module

Solar PV module is otherwise known as solar PV panel. Each unit or module is designed for a rated watt-peak at the standard test conditions (STC). To get the required voltage, current PV modules are wired together to form PV array. The solar module converts sunlight into electrical (DC) energy.

The photovoltaic panels selected have to confine to the updated version of IEC photovoltaic panel requirement check or corresponding BIS standards IEC 61215/IS14286. Also, the panels should match to IEC 61730 Sect. 5.1 (needs for erection) and Sect. 5.2 (needs for checking), for security requirement or corresponding IS. In case of solar panels to be used in extremely chemically reactive surroundings all through its lifespan, it should meet the requirements as per IEC 61701 [3]. The panel casing should be composed of materials that are resistant to corrosion, preferably having anodized aluminium. The permissible variation in output power from the mentioned rated value is $\pm 3\%$. The maximum power extraction voltage value and maximum power extraction current value of every panel or sequence of panels should not change by more than 2% from the corresponding specified value for the panel and from the specified arithmetic average value for the string.

5.2.2 Module Mounting Structures

It is the base on which the modules rest. It should be capable of withstanding rain, hail, wind, adverse climatic conditions, etc. Mounting solar panels correctly is part of maximizing power generation. The proper solar panel mounting provides stability and the proper direction and latitude orientation for the array. Hence care should be taken in robustness of design, incline angle, facing direction, PV array shadowing, etc. High-temperature electroplating of zinc-plated MS frames structures has to be made use of for placing the panels/arrays. Least thickness of galvanization shall not be less than 120 microns. Every structure shall have angle of tilt as per the latitude and location circumstances to get highest solar radiation. On the other hand, to have room for accommodating more number of panels, the direction of leaning has to be decreased until the panels reach the specific index of performance ratio criteria. The frame structure made of steel should, according to newest IS 2062: 1992 and electroplating of the frame structures, be in conformity with latest IS 4759. Frame substance should be anti-corrosive as well as electrolytically well-matched along with the other components made use of in the panel structures like fixtures used for fastening it with the frames [8]. The fixtures used have to be stainless steel material.

The frames have to be made so as to allow simple substitute of any panel. The panel structure should be planned in such a manner that it utilizes least area with no adverse effect on the productivity of solar PV modules [2]. The total mass of the frames (while set up along with PV panels) above the rooftop shall not be more than 60 kg/m^2 . The lowest space between the frame and the rooftop plane shall not be less than 300 mm.

5.2.3 DC Combiner Box (Junction Box)

A DC junction box (JB) is recommended to be utilized for connecting cables from the solar panel to evacuate DC power produced in the panel. It should include DC fuse unit for the leaving DC cable(s) which enters the DC DB. The combiner units have to be made available in the solar array structure for joining of linking cables. Combiner unit should be manufactured using GRP/FRP/powder-coated aluminium/cast aluminium alloy with full dust, water sealed and pests-proof manner. Every conductor shall be ended with cable shoes [3]. Junction boxes should be in such a manner that the incoming and outgoing connections can be done from end to end with appropriate sealing and termination device to make certain that the features of the cable which enters can remain unchanged. Copper bus bars located in the combiner unit with appropriately marked cable ends in compliance with IP65 standard and IEC 62208. Door having EPDM rubber gasket shall be used to avoid water to enter. Single-/double-compression sealing and termination device to make certain that the features of the cable which enters remain unchanged should be made available. Provision of earthing shall be ensured. The provision for earthing shall be positioned at 5 feet or more for convenience of access [7]. Each combiner unit is supposed to have appropriately rated capacity metal oxide varistors/surge protection devices and appropriate reverse blocking diodes. The combiner units should have appropriate provision for keeping the system under observation and detachment options in case needed for every group. Proper identification should be given to each bus bar in order for trouble-free recognition. Every fuse unit has to be kept inside thermoplastic IP 65 casing having see-through covers.

5.2.4 DC Distribution Box (DC DB)

It should be placed near grid-interactive inverter, which is connected to the service mains. The DC DB should be made of thermoplastic IP65. It should include the subsequent apparatus along with cable ends: receiving positive as well as negative DC wires leaving the DC junction unit; DC CB, DC (surge protection device [SPD]), class 2 according to IEC 60364-5-53; and leaving positive as well as negative DC conductors to grid-interactive inverter. In case DC isolator is being connected in place of a DC CB, a DC fuse unit is connected inside DC DB [3].

5.2.5 Power Conditioning Unit (Inverter)

Inverter helps to modify DC power generated by photovoltaic panels into AC power needed for electrical equipment or deliver it to utility grid. The converter is generally designed so as to handle load demands, its switching in and out power transitions and its power quality [5]. Nevertheless, along with this there are other significant criteria for quality inverter choice, like one or more maximum power point tracking (MPPT) units, conversion efficiency, self-power consumption, and power quality.

As solar PV array generates direct current electrical energy, it is essential to change this DC into AC and regulate the voltage levels according to the utility voltage [9]. Power transfer can be brought about using power converter along with accessory control and security equipment. The elements of the system used for transforming the power as per load requirement are referred to as power conditioning unit (PCU). Additionally, it can carry MPPT unit as well, which is couple dam id photovoltaic panel and the converter. The inverter is also required to be interconnected with utility grid/a stable power source. Inverter output shall be well-matched with the grid voltage magnitude and frequency. PCU/inverter should be able to operate entirely by itself without any manual interaction, like starting up, tying up with mains and cut-off [8]. For observing system parameters external computer is required, also the same parameters are recorded in the computer for future analysis.

The PCU should have protection against islanding in compliance to IEEE 1547/UL 1741/IEC 62116 or corresponding BIS standard. In PCU/inverter, there should be a DC cut-out provision made available at the output with the help of an appropriate isolating transformer. If it is not included along with PCU/inverter, an external isolation transformer of appropriate capacity has to be connected externally to the PCU/PCU units for capacity above 100 kW [1]. The power quality problems injected at the point of coupling to the utility services shall be according to the newest CEA (Technical Standards for Connectivity Distribution Generation Resources) guidelines. The units shall meet the terms of relevant IEC/corresponding BIS standard, on account of useful observation and ecological checks as per standard codes IEC 61683/IS 61683 and IEC60068-2 (1,2,14,30)/corresponding BIS standards. The MPPT units ecological check shall meet the requirements as per IEC 60068-2 (1, 2, 14, 30)/equivalent BIS standard. The combiner boxes/casing shall be IP65 (for outdoor)/IP 54 (indoor) and as per IEC 529 stipulation.

5.2.6 AC Distribution Box

It should be positioned nearby the grid-interactive inverter. It should be housed in thermoplastic IP65 frame type as well as it comprises the subsequent devices along with conductor ends: receiving end 3-core/5-core (single-phase/three-phase) cable

from the grid-interactive inverter alternating current CB, 2-pole/4-pole – AC (SPD), class 2 according to IEC 60364-5-53, leaving conductor terminals to the building DB. AC Distribution Panel Board (DPB) should manage the AC power from PCU/inverter and shall have required surge arrestors. Interconnection is required from the DB to mains at low-voltage bus bar during the grid-tied mode. All switchgears and connectors shall match to IEC 60947, part I, II and III/IS 60947 part I, II and III specifications. All the modules shall be metal clad, totally covered, stiff, floor mounted, air-insulated, cubical type suitable for operation on three-phase/single-phase, 415 or 230 volts, 50 Hz. The module should be calculated for lowest likely surrounding temperature of 45 °C, 80% humidity and dusty weather. All indoor modules shall have protection requirement as per IP54 or superior than it. All outdoor modules shall have protection requirement as per IP65 or superior than it. Apparatus and links shall match with Indian Electricity Act and rules. All the 415 AC or 230 volts devices/equipment like bus support insulators, circuit breakers, SPDs, VTs, etc. placed inside the switchgear should be appropriate for constant functioning and acceptable performance during deviation in supply of $\pm 10\%$ and deviation in incoming frequency of ± 3 Hz.

5.2.7 Cables

All conductors should be selected in compliance to IEC 60227/IS 694 and IEC 60502/IS 1554. For the DC cabling, XLPE or XLPO insulated and sheathed, UV-treated single-core flexible copper cables should be used. Multi-core cables should not be selected for this purpose. For the AC cabling, PVC or XLPE insulated and PVC-covered single- or multi-core bendable copper cables should be used. Outdoor AC cables should have a UV-protected external cover. The total voltage drop on the cable section from the solar PV panel to the solar grid-tied inverter should not be more than 2.0%. The total voltage drop on the cable section from the solar grid-tied inverter to the distribution board of the user should not be more than 2.0%. The DC cables from the solar PV panel array should be drawn all the way through a UV-resistant PVC canal pipe of sufficient diameter with a least wall thickness of 1.5 mm. Cables and wires used for the linking of solar PV panels should be used with solar PV connectors (MC4) and couplers. All cables and conduit pipes should be affixed to the rooftop, parapet and ceilings by means of thermoplastic clamps at spacing not greater than 50 cm. The least DC cable size should be 4.0 mm² copper. The smallest AC cable size shall be 4.0 mm² copper. Cables and conduits which have to go by walls or ceilings should be drawn via a PVC pipe sleeve. Cable conductors shall be ended with tinned copper end-ferrules to avoid fraying and breaking of wire strands. The termination of the DC and AC cables at the solar grid-tied inverter should be made as per directions of the company personnel; this in the majority of cases can consist of the use of particular connectors.

5.2.8 Protection: Earthing, Lightning and Grid Islanding

The arrangement shall be made available with all essential protections like earthing, lightning and grid is landing as follows:

Earthing Protection Every array unit of the PV work area shall be grounded/earthed appropriately as per IS:3043-1987. Additionally the lightning arrester/masts shall also be earthed within the array region. Earth resistance should be checked as and when necessary following earthing by calibrated earth tester. PCU, AC Distribution Board and DC Distribution Board shall also be earthed appropriately. Earth resistance should be less than 5 ohms. It should be made sure that all the earthing points are connected mutually to bring them to the identical potential level.

Lightning Protection The solar PV power plants should be equipped with lightning arrestors and overvoltage relays. The chief goal in this regard should be to decrease the excess voltage to a reasonable value before it reaches the PV or other subsystem components. The cause of high voltage could be lightning, faults etc. The whole space occupying the solar PV array should be properly secured against lightning by using necessary number of lightning arrestors. Lightning arrestors shall be used as per IEC 62305 standard [11]. The safeguard against induced excess voltages should be given by the use of metal oxide varistors (MOVs) with proper earthing so as to provide transient surges as safe path to earth.

Surge Protection Interior surge security should comprise of three MOV-type surge-arrestors each connected from positive and negative terminals to earth (via star arrangement). Surge arrester should be connected on the DC side and the AC side of the solar system configuration. The DC surge protection devices (SPDs) should be connected in the DC DB nearby to the solar grid-tied inverter. The AC SPDs should be connected in the AC DB adjacent to the solar grid-tied inverter [2]. The SPDs earthing terminal should be provided to earth via the before described dedicated earthing arrangement. The SPDs should be of type 2 as per IEC 60364-5-53.

Grid Islanding In the case of a power outage on the utility grid, it is necessary that any autonomous power-producing inverters connected to the service mails should cut out in a small span of period. This avoids the DC-to-AC inverters from continuing to supply power into local portions of the service mains, known as "Islands." Anti-islanding protection in compliance to IEEE 1547/UL 1741/ IEC 62116 or corresponding BIS standard should be connected. Electrically live islands impose a risk to labourers who are likely to anticipate the section to be unpowered, and they may also damage grid-tied devices. The rooftop PV system should be provided with islanding safety. Along with detachment from the utility (owing to islanding protection) detachment owing to low- and high-voltage circumstances should also be given.

5.3 Sizing of Solar PV System

Grid-interactive photovoltaic arrangement supplies power straight to consumer appliances without intermediate storage. Excess energy, if available, is delivered to the service mains and deficit, if at all, is consumed from the service mains. To assist such transfer of power, a provision recognized as net metering has been put in place [4]. In net metering, the solar energy delivered to the service mains is subtracted from the energy consumed from the service mains. The user pays for the net-energy consumed from the service mains. To facilitate net metering, electricity service providers will substitute the actively available utility connection meter with a bidirectional meter that shows and monitors the consumed and delivered energy independently [9].

5.3.1 Solar Array Capacity

The shadow-free area necessary for setting up of a rooftop solar PV system is around 12 m² per kW (kilowatt). This figure consists of prerequisite of space amid array rows. The solar module has been set up on the top of the structure with a south oriented incline angle that changes within Tamil Nadu from 11 to 13 degrees based on latitude position of the site. Adequate space should be made available for maintenance of the structure. The least amount clearance is needed for maintenance with cleaning of the module is 0.6 m from the fortifications wall as well as in the middle of rows of modules. Among the rows of arrays, adequate space is to be left to avoid the shadowing of a row by another nearby row.

In Tamil Nadu the usual mean yearly energy production per installed kW of photovoltaic system is 1500 kWh. (This figure is obtained by considering a utilization factor of 18% and mean grid accessibility of 95% throughout the hours of daylight.) According to Tamil Nadu solar net-metering rule, electricity produced by a solar rooftop which is exported into the TANGEDCO service mains should be limited commercially at 90% of the electricity utilization by the appropriate user at the end of a 12-month agreement duration which is from August to July. Surplus power produced above the 90% limit should be considered as lapsed and would not be attributed for the consideration of net metering.

The suggested capacity of the solar panel array of a planned grid-interactive solar PV system can be designed by means of the following steps:

First: Determine the utmost system capacity depending on the shade-free rooftop area available.

$$C = A/12,$$

where C is the capacity required and A is the available shadow less rooftop region (in square meters).

Assume the shadow less rooftop region is 60 m². The highest solar PV capacity that can be set up on this rooftop space could be 5 kWp

Secondly: Determine the system capacity depending on yearly energy utilization.

$$C = (0.9 \times E)/1500$$

where E = Yearly energy utilization (in kWh)

Assume the yearly energy utilization is 15,000 kWh. The solar PV system capacity depending on yearly energy utilization could be 9 kWp.

Finally capacity to be chosen is the lowest value obtained from the above two methods.

In this illustration: 5 kWp.

5.3.2 Inverter Sizing

The suggested grid-interactive inverter rating in kW can be in the order of 95–110% of the photovoltaic panel capacity. As per the given illustration mentioned in the previous section, the photovoltaic panel aggregate rating was designed to be 5 kW [9]. The grid-interactive inverter needed for the above system could of the order of 4.75–5.50 kW. The type of load has a determining factor when sizing of inverter is concerned as it has to withstand the surges at the time of starting. The inverter capacity has to be 25–30% higher than total power requirement of connected equipment. If the connected equipment is motor or compressor load type of equipment, then inverter size should be almost three times higher than the connected equipment rating.

The aggregate power of every inverter connected in the system shall be more than installed capacity of solar power plant for STC. MPPT unit shall be incorporated in the PCU/inverter to maximize utilization of energy generated by the panel.

5.4 Solar PV–Grid Interfacing

The interfacing of solar PV system with grid is done via the grid-interactive inverter. The power produced from solar PV array is provided to the converters that transform DC generated by solar PV panel to AC and supplies to utility grid after for power evacuation. If there happens to be a grid outage or grid faults leading to large-voltage fluctuations, converter will be cut off from the utility grid. Only after the utility service is restoration will the system start yet again to tie up with utility mains, and user demand would be supplied to the degree possible depending on the availability of power.

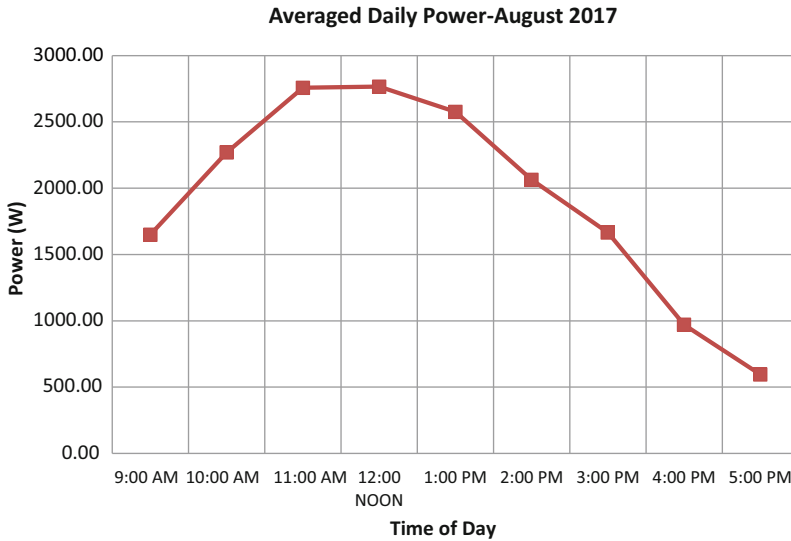


Fig. 5.2 Daily averaged power for the month of August

5.5 Results and Discussion

The averaged diurnal variation was plotted as in Fig. 5.2. The same procedure was followed for all the months for a year from September 2016 to August 2017.

On an average the daily power production for the month of August is 1923.57 W, and its variation over the day is given in Fig. 5.3. Hence the daily energy output can be calculated as product of daily average power and number of sun hours during the day, which is 8 h in the above case. After taking monthly plot, yearly output was obtained. The daily peak, the average daily peak and the monthly peak were obtained. Checking the maximum values and the mean maximum value over the month can help to ensure the system capacity. And hence its ability to withstand the peak production in any such event is assured.

5.6 Conclusion

As the PV industry is quickly advancing, the pace by which modern, highly sophisticated cutting-edge technology is evolving is highly remarkable. However, so as to make use of the expertise and options available in a resourceful manner and to take full advantage of the options available, proper consideration of assortment along with proper measures and specification of the BoS elements becomes necessary. Appropriate choice helps in reducing losses in the range of 10–20%, whereas

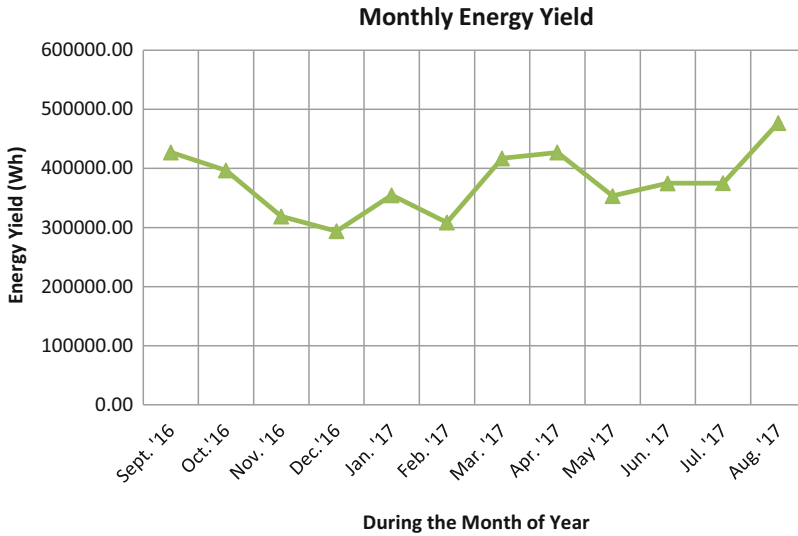


Fig. 5.3 Monthly energy output

improper selection might simply decrease the system effectiveness by 50% or less. So, it is essential to make sure that sufficient care is given while choosing each and every BoS element.

The design of a 5 kW solar PV power plant placed on the rooftop of PSG College of Technology, Coimbatore, is carried out by way of specifying the engineering principles and practical constraints of site. Detailed discussion on how to establish 5 KW photovoltaic solar rooftop power plant design as well as power production calculations is done. Special care is taken to adhere to standards for each component selection. Special care is taken on selecting BoS components which is normally neglected but actually accounts for more than 50% of cost of solar PV system installation and maintenance.

References

1. Roger A Messenger, and Jerry Ventre, Photovoltaic systems engineering, 2nd edn. CRC Press, Boca Raton
2. Yang Y, Ye Q, Tung L, Greenleaf M, Li H (2017) Integrated size and energy management design of battery storage to enhance grid integration of large-scale PV power plants. *IEEE Trans Ind Electron* 65(99):1-9
3. "A Guide to Photovoltaic (PV) System Design and Installation", Prepared by Endecon Engineering, Prepared for California Energy Commission, Energy Technology Development Division, June 2001
4. Raj A, Gupta M, Panda S Design simulation and performance assessment of yield and loss forecasting for 100 KWp grid connected solar PV system. 2nd IEEE Int Conf on Next Generation Computing Technologies (NGCT), pp 528-533, October 2016

5. Dubey K, Shah MT (2016) Design and simulation of Solar PV system. IEEE International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), pp 568–573
6. Radhey SM, Rathore JS, Shivani J (2014) Grid connected roof top solar power generation: a review. Int J Eng Dev Res 3(1):325–330
7. Zapata JW, Perez MA, Kouro S, AnssiLensu AS (2015) Design of a cleaning program for a PV plant based on analysis of energy losses. IEEE J Photovoltaics 5(6):1748–1756
8. Malla A, Niraula A (2012) Importance of balance of system in a solar PV application. Rentech Symp Compendium 2:65–69
9. “Guidelines for Grid Connected Small Scale (Rooftop) Solar PV Systems for Tamil Nadu”, Version: April 2014
10. Youngren E Why whole system integration and balance of system components are crucial to off-grid PV system sustainability. 2011 IEEE global humanitarian technology conference
11. Ganguli S, Sinha S (2011) Design of a 11kWp Grid connected solar photovoltaic power plant on 100m² available area in the Birbhum District of West Bengal. J Inst Eng 8(1):105–112
12. Patel MR Wind and solar power systems design, analysis and operation, 2nd edn. CRC Press/Taylor and Francis Group, Boca Raton
13. Sumathi S, Ashok Kumar L, Surekha P Book on “Solar PV and wind energy conversion systems – an introduction to theory, modeling with Matlab/Simulink, and the role of soft computing techniques” – *Green energy and technology*, Springer; 2015 edition (20 April 2015), ISBN-10: 3319149407, ISBN-13: 978-3319149400
14. Sumathi S, Ashok Kumar L, Surekha P (2015) Book on “computational intelligence paradigm for optimization problems using MATLAB/SIMULINK”. CRC Press/Taylor and Francis Group, Boca Raton
15. Sharma VP, Singh A, Sharma J, Raj A Design and simulation of dependence of manufacturing technology and tilt orientation for 100kWp grid tied solar PV system at Jaipur. IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), pp 1–7, December 2016

Chapter 6

An Add-in Tool for BIM-Based Electrical Load Forecast for Multi-building Microgrid Design



Jasim Farooq, Rupendra Kumar Pachauri, R. Sreerama Kumar, and Paawan Sharma

Abstract For effective multi-building smart microgrid design and optimization, it is necessary to gather sufficient and accurate computational data that too at the early design stage of projects. The building level electric energy demand forecast is one of the significant steps for suitable generation planning and to formulate strategies for demand response. The adoption of information and communication technologies (ICT) in construction such as building information modelling (BIM) added significant values by improving productivity and by providing digital and object-oriented data-rich realistic integrated models for engineering calculations and coordination. This paper discusses the development of an add-in tool as an Autodesk Revit 2017 plug-in application for the BIM-based load forecast estimation to be used for the electrical design of multi-building microgrids. The calculation proceeds by summing up the individual forecasts of each elementary load components by following a simplified bottom-up approach. The proposed tool is capable of generating realistic electrical load profile for a selected time period on an hourly basis.

Keywords BIM · C# · Load forecasting · Revit · Smart microgrid

6.1 Introduction

Smart microgrid (SMG) is one of the effective technologies for achieving reliable and secure electrical supply by combining smart devices and algorithms, modern communication technologies and utilization of distributed renewable energy resources.

J. Farooq (✉) · R. K. Pachauri · P. Sharma
School of Engineering, University of Petroleum & Energy Studies, Dehradun, India

R. Sreerama Kumar
Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah, Saudi Arabia

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_6

SMG system design optimization is a complex process with numerous constraints that have to be taken into consideration such as engineering, environmental, financial and project specifications. For optimal planning of SMG, one of the requirements is the accurate forecasting of hourly demand load, monthly peak loads and details of the load factor of the proposed construction projects. Underestimation or overestimation of electrical energy requirements on certain time may lead to additional costs, revenue loss and ineffective operation of SMG. The load demands influenced by many factors such as ambient weather and occupant's behavioural and socioeconomic factors [1] lead to a time variant and non-linear function. Precise load forecasting is important to provide necessary information for the optimum sizing of power system components and operations of the electrical infrastructure. Different types of load forecasting methods are existing and bottom-up approach presented in [2, 3] that is based on individual load's mean daily starting frequency (MDF), mean consumption cycle time length (MCCT) and starting probability (SP) is used to develop the proposed building information modelling (BIM)-based add-in tool named building information modelling load profile estimator (BIM-LPE). Detailed bottom-up load forecasting is a time-consuming and tedious work. Experienced forecasters having ample understanding are required for load forecasting. As redundant calculations are to be done, technologies that can precisely give results without or with least user input are preferred. Emerging BIM can provide faster computational platform and gather various kinds of load forecasting data such as electrical load-type details and behavioural patterns. BIM is the process of developing and managing the N th-dimensional integrated and realistic depiction backed by an information system of our construction requirements. BIM adoption for MEP design will be a key approach for attaining energy efficiency in buildings. From BIM models, load forecasting estimation and energy audits by considering elementary device load profiles and occupant's behaviours are possible by developing BIM customized tools.

Autodesk Revit and Graphisoft's ArchiCAD are leading matured BIM applications and Revit Application Programming Interface (API) permits developers to extend the capabilities by writing the scripts using .NET languages such as C# and Visual Basic that adds new functionality. The Revit API consents programmers to access elements in the BIM to retrieve and modify the parameters to conduct specialized tasks such as load forecasting and behavioural analysis. Once BIM-authored tools provide digital details of load profile data such as types of loads, hourly consumption and activities data, then with the help of Geographical Information System (GIS) technologies, normalized regional load profile data can be extrapolated. As the wide variety of buildings are existing, the regional-level typical average energy consumption forecasting requires a large database and computational requirements and the same can be achieved by BIM and GIS integration. This paper describes the development of an Autodesk Revit 2017 plug-in application for the BIM-based load profile forecast estimation of buildings. The calculation proceeds by summing up the individual forecasts of each elementary load components by following a simplified bottom-up approach. The remaining parts of this paper are structured as follows: Sect. 6.2 discusses the literature review and Sect. 6.3 describes the methodology. Section 6.4 presents the case study and paper conclude by Sect. 6.5.

6.2 Literature Review

For the new construction projects with SMG, distributed energy source, technology and size selection and the concept of SMG planning such as Demand Side Management (DSM) require thorough information of timely varying electrical energy pre-occupancy load forecasting [4]. Detailed information of household load curve is required for analysing the impact of the photovoltaic and electric vehicles for preparing strategies like DSM [5]. Forecasting and estimating energy demand profiles for domestic projects are a well-studied area. In the literature, many models are proposed for load profile forecasting and they can be broadly classified into top-down and bottom-up approaches [6, 7]. A typical drawback for accurate bottom-up methods is the huge data requirements about the consumer's behaviours and their building's electrical individual loads and mostly, some data are not easily obtainable [2]. A complete load forecasting procedure often involves complex and huge data collections, sophisticated software with seamless interoperability, advanced mathematical modelling and accurate documentation process. Load forecasting is a dynamic process and authorities should be regularly looking for methods that can advance the process and the databases. It is vital that specialists dedicate significant time and funds to develop reliable load forecasts [8]. One of the promising methods to solve above discussed requirements is by using information and communication technologies (ICT) for construction such as Building Information Modelling (BIM) and software developers can combine existing huge collection of measured data related to energy use with building modelling to achieve more realistic energy models [9]. BIM capabilities and benefits are well studied and validated. The major benefits of BIM are lowering capital project costs, better design, coordination, communication, integrated engineering analysis, visualization and project pre- and postconstruction management [10]. Authentic researches are carried out related to the application of BIM for automated code checking, costing and load estimation [11]. Even though BIM offers a plethora of possibilities for electrical system developments, historically BIM-based electrical studies and deployment are lagging behind other trades such as civil and mechanical [12]. Shen et al. [13] carried out BIM-enabled activity simulation and [14] concludes that inhabitant behaviour modelling can be combined with BIM. BIM-based real-time monitoring of electrical consumption was discussed by [15] for the reduction in energy consumption and [16] presented BIM-supported real-time building energy modelling [17]. [17] discussed BIM-enabled real-time power consumption visualization [18]. [18] presented an ANN- and BIM-based methodology for forecasting the energy price of residential houses in Saudi Arabia. BIM-enabled pre-occupancy estimation of load profile can be possible by combining load forecasting techniques, historical (locally or web-stored) data and BIM-based automated tools for information retrieving and processing. BIM-based load profile estimation along with building load details can be exported to GIS for macro-scale analysis as various methods exist for BIM and GIS data sharing and for seamless integration [19]. By scaling micro-level BIM comprehensive data to macro-level data in GIS, one can generate representative

characteristics of each building archetypes by normalizing consumption data as done in [20] and by utilizing technologies like big data analytics. A web-based GIS interactive system can effectively store individual load profile data and reproduce when it is required. BIM and GIS combined city energy information modelling by considering load profile of each building along with detailed load information can effectively narrow down SMG planning issues. As per authors' understanding, studies related to BIM-enabled load profile studies are limited and in this paper a novel BIM-based tool for load profile estimation is presented.

6.3 Methodology

As per authors' understanding, studies related to BIM-enabled load profile studies are limited, and in this paper a novel BIM-based tool for load profile estimation is presented. The bottom-up load forecasting approach presented in [2, 3] is the base to develop BIM-LPE and this is capable to identify the impact of each end-use electrical load towards the collective electrical energy consumption of the building. By assuming SP and MCCT are the same for both weekday and weekend day throughout the year, then

ON Time of an individual appliance in an hour (OT_{ah}) in minutes =

$$SP_{ah} * MDF_a * MCCT_a / 100 \quad (6.1)$$

Hourly energy consumption by an appliance (kWh_{ah}) =

$$W_a * OT_{ah} / (10^3 \times 60) \quad (6.2)$$

Total monthly energy consumption =

$$30 \left(\sum_{(a=1)}^n \sum_{(h=1)}^{24} kWh_{ah} \right) \quad (6.3)$$

Where

SP_{ah} = hourly starting probability of an appliance.

MDF_a = Mean daily starting frequency of an appliance.

W_a = wattage of an individual appliance.

kWh_{ah} = kWh consumption of an individual appliances in an hour.

n = total number of appliances and other electrical loads.

BIM-LPE's coding is done by following Revit API Developers Guide [21]. BIM-LPE developing has three steps and they are explained below.

Main Programme As per the flowchart shown in Fig. 6.1, coding is done in Microsoft visual studio C#. Eqs. (6.2) and (6.3) are used in the main programming

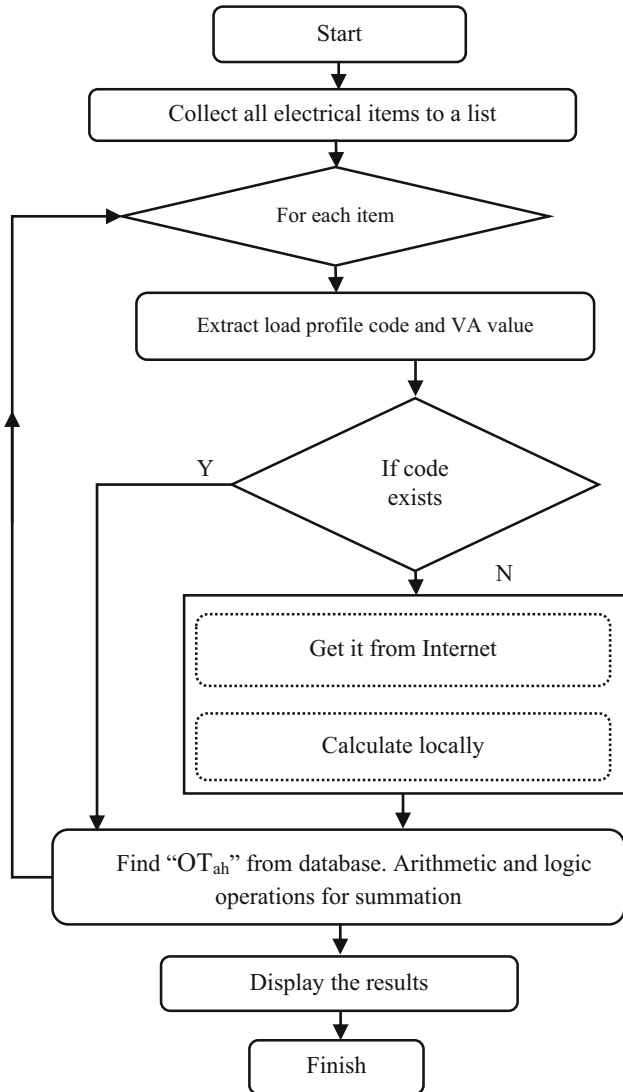


Fig. 6.1 Flowchart of main programming

to estimate load profile data. By using windows forms, user interfaces are developed as shown in Figs. 6.4 and 6.5 and encapsulated the functionality of BIM-LPE in them. For storing individual load profile data locally, SQL is used and necessary coding is done to retrieve and store. Each family of BIM model is to carry a load profile code (LPC) and BIM-LPE will find previously stored OT_{ah} data from SQL database according to LPC. The concept of generating LPC, which is used in BIM-LPE, is given in Table 6.1. Auto-generation of LPC and OT_{ah} for every

Table 6.1 Load profile coding (LPC) example for the microwave oven having LPC ‘SA-20146-8124-1BHK_R-MW’

Description	Code
Country code	SA
ZIP code	20146
Building Number	8124
Archetype	1BHK_R
Load type	MW



Fig. 6.2 Created BIM-LPE add-in button in Revit

electrical element by BIM-LPE with the help of programme-specific ontologies and web-based data is to be done in future work.

Button Programme A new button is created in Revit by coding button’s icon and main programme’s path details as per Autodesk guidelines [21] and pressing that button invokes BIM-LPE. Once the coding and registering of the BIM-LPE are done, then BIM-LPE will automatically be invoked when Revit starts up and will appear on the add-ins tab as encircled in Fig. 6.2.

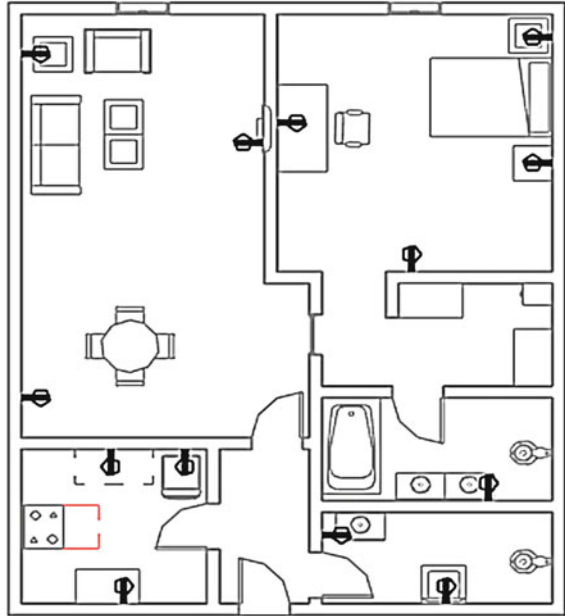
Manifest File BIM-LPE is registered with Revit application by way of an .addin manifest file and is to be placed at specific folders as per [21]. The programmes provided by Autodesk [21] are used to develop the manifest file.

6.4 Case Study and Inferences

A case study is developed in Revit 2017 platform for a 1 BHK residential apartment as shown in Fig. 6.3. Each electrical element defined with LPC and related ‘OT_{ah}’ as shown in Table 6.2, are added to SQL database by using BIM-LPE’s functionality. The generated load curve is shown in Fig. 6.4 and the user interface for adding ‘OT_{ah}’ values to SQL database is shown in Fig. 6.5. The runtime for BIM-LPE is around 2 s.

Evolving and dynamic taxonomy of BIM electrical families with appropriate energy consumption details is to be brought to standardize load profile estimation. BIM-LPE can assist for cost savings, finalizing distributed energy generation mix selection and smart algorithms for DSM while planning a SMG project. Each

Fig. 6.3 Plan view of developed case study



building load profile design data can be stored and normalized by using GIS tools and can systematically develop the most suitable SP, MCCT and MDF for load types according to different building archetypes and house types as explained in Fig. 6.6. Continuously varying load profile data normalization is possible by the seamless integration between BIM and GIS and that leads to precise comprehensive national electrical energy consumption model.

Pre-construction design BIM can be used to develop real-time power consumption modelling which can exactly locate when, where and how power consumption takes place in the building. So from both pre- and postconstruction BIM, different kinds of qualitative and quantitative load profile data can be transferred to GIS. By analysing and normalizing each and every substantially important data, policies for energy conservation and smart built environment can be drawn by authorities and can generalize load pattern forecast for each archetype at the provincial level. National and regional average inaccuracies can be narrowed down by combining monthly billing, sub-metered data details and BIM data (detailed housing information) and geographical and social conditions. This method can minimize detailed surveys and can reduce the time for analysis. To implement BIM-LPE or similar BIM-based tools, faster adoption of BIM by electrical contractors and standardization of modelling is a necessity. The amount and kind of data extracted from BIM for GIS level analysis require extensive studies for industrial applications.

Table 6.2 Values of ‘OT_{ah}’ in minutes for different load types

	MW	REF	DW	WM	TV	PC	GP
1	0.021	20.26	0.104	0.027	3.978	1.428	0.107
2	0.003	20.26	0.000	0.000	2.269	0.815	0.056
3	0.000	20.26	0.000	0.000	1.017	0.365	0.055
4	0.000	20.26	0.000	0.000	0.901	0.323	0.051
5	0.000	20.26	0.000	0.000	1.018	0.365	0.056
6	0.010	20.26	0.000	0.000	1.134	0.407	0.073
7	0.101	20.26	0.000	0.000	1.134	0.407	0.089
8	0.155	20.26	0.146	0.039	1.708	0.613	0.149
9	0.256	20.26	0.417	0.112	2.843	1.021	0.171
10	0.349	20.26	0.962	0.257	3.978	1.428	0.168
11	0.409	20.26	1.465	0.391	4.540	1.630	0.158
12	0.462	20.26	1.509	0.403	5.674	2.037	0.166
13	0.465	20.26	1.509	0.403	5.674	2.037	0.171
14	0.420	20.26	1.532	0.409	6.938	2.490	0.187
15	0.375	20.26	1.532	0.409	7.172	2.574	0.208
16	0.346	20.26	1.532	0.409	7.956	2.856	0.252
17	0.398	20.26	1.551	0.414	7.956	2.856	0.265
18	0.435	20.26	1.551	0.414	7.956	2.856	0.287
19	0.430	20.26	1.616	0.431	9.090	3.263	0.308
20	0.425	20.26	1.616	0.431	9.652	3.465	0.317
21	0.407	20.26	1.551	0.414	7.956	2.856	0.285
22	0.240	20.26	1.277	0.341	6.247	2.243	0.280
23	0.135	20.26	0.816	0.218	5.674	2.037	0.203
24	0.059	20.26	0.187	0.050	4.539	1.630	0.135

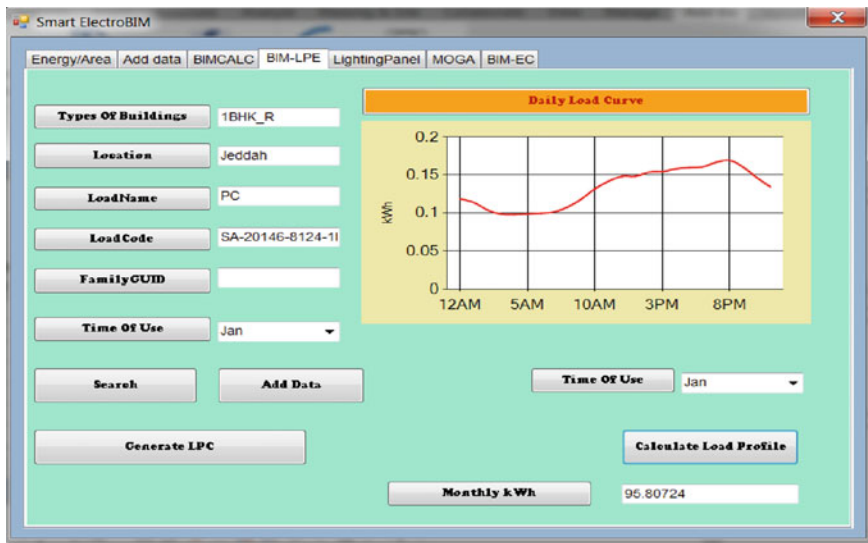


Fig. 6.4 The BIM-LPE main user interface with the results of a case study

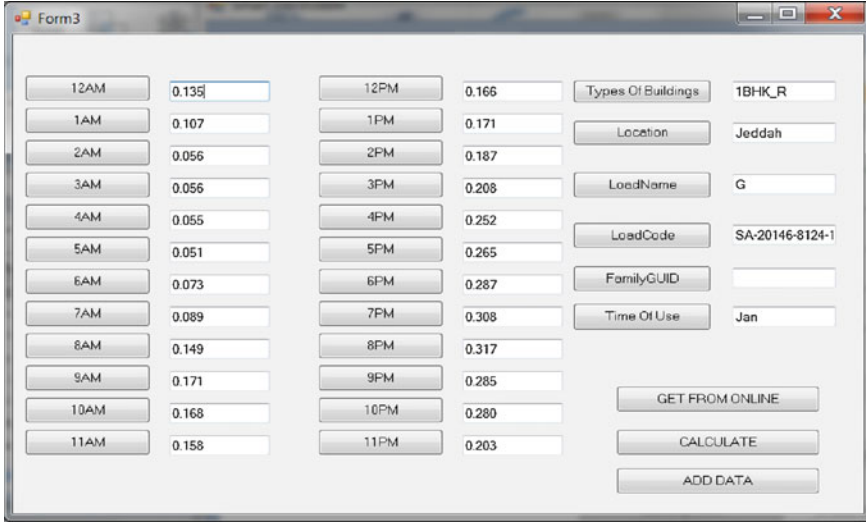
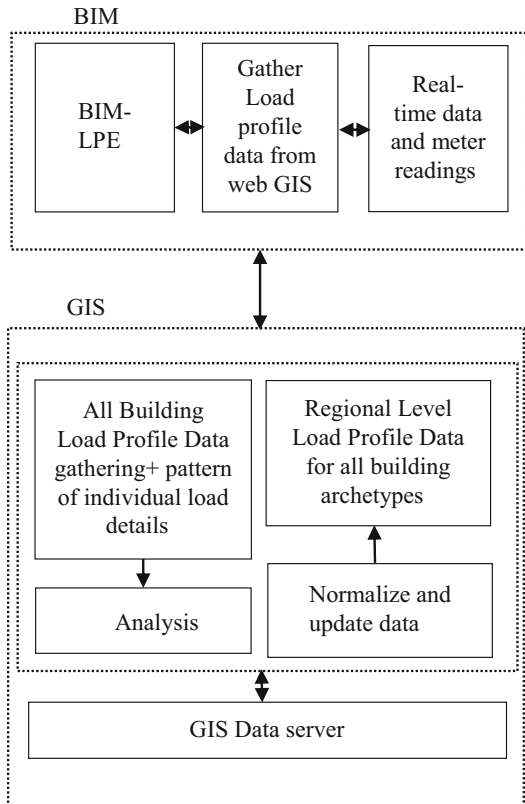


Fig. 6.5 The user interface for adding 'OT_{ah}' values to SQL database

Fig. 6.6 BIM and GIS combined load profile data standardization



6.5 Conclusion

In this paper, we presented a simplified and easy-to-use BIM-based tool for electricity load profile estimation of buildings for a specified time period. A case study was conducted according to BIM-LPE and results are presented. BIM adoption by the industry and standardization of modelling is required for efficient BIM-authored tools such as BIM-LPE, and government construction strategies are to address the faster acceptance of BIM by the public. Integrated pre- and postoccupancy BIM design and analysis data scaled to GIS will be a cost-effective efficient method for regional-level load profile investigations. As BIM and GIS technologies are becoming mandatory, BIM-LPE and similar tool's applicability will be high. In our future works, improvements of BIM-LPE will focus on developing automated detection ontologies of electrical load types and automated estimation of SP and LPC by improving the intelligence and considering behaviour and activities of consumers.

References

1. Hahn H, Meyer-Nieberg S, Pickl S (2009) Electric load forecasting methods: tools for decision making. *Eur J Oper Res* 199(3):902–907
2. Chuan L, Ukil A (2015) Modeling and validation of electrical load profiling in residential buildings in Singapore. *IEEE Trans Power Syst* 30(5):2800–2809
3. Paatero J, Lund P (2006) A model for generating household electricity load profiles. *Int J Energy Res* 30(5):273–290
4. Hoogsteen G, Molderink A, Hurink J, Smit G (2016) Generation of flexible domestic load profiles to evaluate demand side management approaches. 2016 IEEE International Energy Conference (ENERGYCON)
5. Nijhuis M, Gibescu M, Cobben J (2016) Bottom-up Markov chain Monte Carlo approach for scenario based residential load modelling with publicly available data. *Energ Buildings* 112:121–129
6. Colson C, Nehrir M (2009) An alternative method to load modeling for obtaining end-use load profiles. 41st North American power symposium
7. Train K (1992) An assessment of the accuracy of statistically adjusted engineering (SAE) models of end-use load curves. *Energy* 17(7):713–723
8. Hong T, Shahidehpour M (2015) Load forecasting case study. [ebook] EISPC and NARUC. Available at: <https://pubs.naruc.org/pub.cfm?id=536E10A7-2354-D714-5191-A8AAFE45D626>. Accessed 28 June 2018
9. Kamilaris A, Kalluri B, Kondepudi S, Kwok Wai T (2014) A literature survey on measuring energy usage for miscellaneous electric loads in offices and commercial buildings. *Renew Sust Energ Rev* 34:536–550
10. Eastman C, Sacks R, Teicholz P, Liston K (2013) *Bim handbook*. Wiley, Hoboken
11. Farooq J, Sharma P, Sreerama Kumar R (2018) A BIM-based detailed electrical load estimation, costing and code checking. [online] *iaescore.com*. Available at: <http://www.iaescore.com/journals/index.php/IJECE/article/view/11525>. Accessed 16 June 2018. (forthcoming)
12. Farooq J, Sharma P, Sreerama Kumar R (2017) Applications of building information modeling in electrical systems design. *J Eng Sci Technol Rev* 10(6):119–128

13. Shen W, Shen Q, Sun Q (2012) Building information modeling-based user activity simulation and evaluation method for improving designer–user communications. *Autom Constr* 21:148–160
14. Sun K, Hong T (2017) A framework for quantifying the impact of occupant behavior on energy savings of energy conservation measures. *Energ Buildings* 146:383–396
15. Alahmad M, Nader W, Brumbaugh A, Cho Y, Ci S, Sharif H, Shi J, Neal J (2011) The “BIM’s 4D+” dimension: real time energy monitoring. 2011 IEEE GCC conference and exhibition (GCC)
16. Dong B, Neill Z, Li Z, Luo Dand Shashanka M (2012) An integrated infrastructure for real time building energy modelling and fault detection and diagnostics. In: 5th National Conference of IBPSA-USA
17. Chiang C, Chu C, Chou C (2015) BIM-enabled power consumption data management platform for rendering and analysis of energy usage patterns. *Proc Eng* 118:554–562
18. Alshibani A, Alshamrani O (2017) ANN/BIM-based model for predicting the energy cost of residential buildings in Saudi Arabia. *J Taibah Univ Sci* 11(6):1317–1329
19. Liu X, Wang X, Wright G, Cheng J, Li X, Liu R (2017) A state-of-the-art review on the integration of building information modeling (BIM) and geographic information system (GIS). *ISPRS Int J Geo Inf* 6(2):53
20. Nishio K, Asano H (2006) A residential end-use demand model for analyzing the energy conservation potential of new energy efficient technologies
21. Revit API Developers Guide, Help.autodesk.com (2018) [online]. Available at: <http://help.autodesk.com/view/RVT/2017/ENU/?guid=GUID-F0A122E0-E556-4D0D-9D0F-7E72A9315A42>. Accessed 16 June 2018

Chapter 7

An Investigation on Torque Ripple Minimization of Switched Reluctance Motor Using Different Power Converter Topologies Using Intelligent Techniques



M. Gengaraj, L. Kalaivani, K. Cherma Jeya, P. Eswari Prabha, A. M. Kirthika, and M. Vavuniya

Abstract In the modern era, the fast necessities of the electrical drive are the indispensable part of the industries. This paper describes the modelling, analysis and study of various converter topologies fed switched reluctance motor (SRM) along with the speed control which is going to be achieved by the fuzzy logic controller. Now a days SRM is gaining more and more attention in recent high-speed industrial application due to its simplicity and ruggedness. Due to the double saliency nature of SRM drive, the torque pulsations are the vital challenge as the torque pulsations are relatively higher when compared to other conventional machines. The origin behind the high torque ripple in SRM is transferring the torque production from incoming phase to outgoing phase during the phase commutation period. Due to the presence of higher torque ripple, acoustic noise and oscillations are induced in the torque. To minimize the torque ripples, the current modulation technique plays the vital role which excites the stator phase sequentially. This paper proposes the different types of converter topologies such as asymmetric, C-dump and R-dump which are used to excite the stator phases of SRM with low torque ripples. Here the speed control can be achieved by the intelligent technique like fuzzy logic controller. The converters are to be designed using MATLAB/Simulink platform, and the mean values of torque and speed are to be compared and analysed. From the comparison, a proper converter is chosen for the SRM to maintain the minimum torque ripple.

Keywords Switched reluctance motor (SRM) · Power converters · Torque ripple · Speed control

M. Gengaraj (✉) · L. Kalaivani · K. Cherma Jeya · P. Eswari Prabha · A. M. Kirthika · M. Vavuniya
Department of Electrical and Electronics Engineering, National Engineering College,
Kovilpatti, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_7

Abbreviations

MATLAB	MATrix LABoratory
PWM	pulse width modulation
RPM	rotations per minute
SRM	switched reluctance motor

7.1 Introduction

In a current scenario of advanced modern control and power electronics, switched reluctance motor (SRM) drive has more enthusiasm in high-speed industrial drive applications. Due to its simple structure and ease operation without magnets and brushes, SRMs provide economical production and operation costs. In addition to that, for a wide speed range of control, SRMs have high power output and system efficiency [1]. These advantages will lead the SRM in more attention for several decades. SRM drive needs a position sensor because the individual phase excitations should be synchronized with the rotor position [2].

The existence of torque ripples is due to the torque dip between two subsequent phases excitation. The radial forces are the vital source which leads to the noise produced in SRM [3]. The torque ripple and acoustic noise are not inevitably harmful to the system in all cases, but it depends on the application and requirement. The suppression of torque ripple with adequate control strategy is essential in case of servo control applications or where the even control is required at low speed [4].

The current commutation technique from incoming phase to outgoing phase should be precisely controlled to obtain superior performance as desired torque and stable speed from the SRM drive. SRM torque ripple can be minimized by controlling the supply voltage, current and load along with the turn-on and turn-off angles [5].

Various control strategies are described to the converter of the SRM in accordance to reduce the torque ripples from SRM, by trailing the rotor position alongside [6]. The triggering pulses are afforded for the switches to the converter to excite each stator phase with corresponding to the rotor position [7]. For these conditions, two schemes are available to achieve the control technique: one is to vary the speed of the motor using PWM strategy and other one is to standardize the firing pulse as a function of motor speed.

Since the torque production mechanism is not depending upon the current excitation polarity, the motor drive required only one switch per phase winding. This is the only difference between the conventional AC drives whereas there minimum two switches are needed for the current control. All the converter topologies such as asymmetric bridge, C-dump, R-dump and bifilar type converters assume that a DC voltage source is available for their inputs [8].

Since SRM has a highly non-linear characteristic, it is required to introduce the intelligent control techniques like fuzzy logic and neural networks methods for its control, and the recent research have more attention on dynamic control of SRM drives using these artificial intelligence control techniques [9]. There are numerous intelligent control techniques adopted and implemented to control the speed of the drive which is used for the industrial applications along with the minimized torque ripple.

For SRM to operate, switching of phases of SRM is important. Many converter topologies were available for the said switching operation. Some converters such as an asymmetrical converter, R-dump converter and C-dump converter for SRM are discussed in this paper. The torque of SRM can be controlled by controlling the current excitation of the stator phase and by the way the torque ripples are the main limitation. This paper depicts the SRM drive fed different converter topologies along with their current control strategy. A comparative study of speed control of SRM with closed-loop control for different types of converters is presented in this paper [10].

The remainder of the paper is organized in the following way. Section 7.2 represents the necessity of torque ripple minimization of SRM. Section 7.3 represents the different types of power converter fed SRM drive in MATLAB environment, and Sect. 7.4 seeks the results of the proposed model and its performance parameter comparison. Section 7.5 concludes the paper.

7.2 SRM Drive Modelling

7.2.1 Configuration

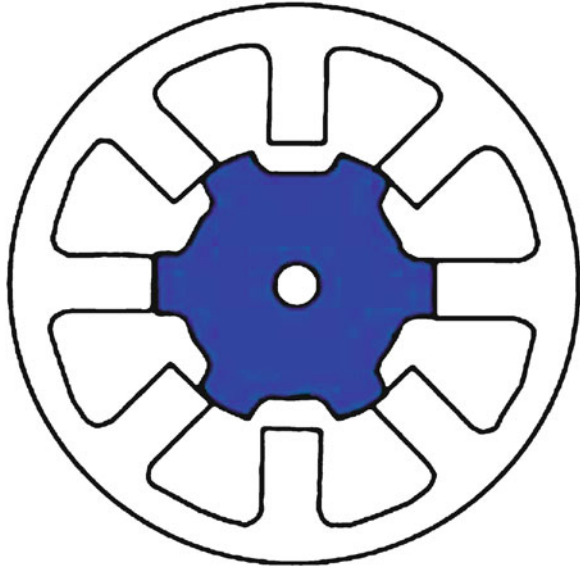
Unlike common DC motor types, power is delivered to windings in the stator rather than the rotor. SRM is a doubly salient, singly excited motor. Stator and rotor are made up of silicon steel stampings [11]. In rotor, poles are projected outwards as shown in Fig. 7.1.

Although the machine has a simple structure, the behaviour of its electromagnetic is convivial [12].

7.2.2 Torque Production Mechanism

The reluctance principle for torque production is utilized in these machines, where the phases operate independently and in succession. The magnetization pattern of the individual phases together with the $T-i-\theta$ characteristics of the motor dictates the amount of torque ripple during operation one method is to improve the magnetic design of the motor, while the other method is to use sophisticated electronic control [13].

Fig. 7.1 Switched reluctance motor drive



One of the performance measures of SRM is torque ripple coefficient (T_i) which is expressed as follows (7.1):

$$T_i = \frac{T_{\max} - T_{\min}}{T_{\text{mean}}} \quad (7.1)$$

where T_{\max} , T_{\min} and T_{mean} are the maximum, minimum and mean values of the total torque, respectively.

7.3 Modelling of SRM Drive with Various Power Converters

As the torque is independent of the direction of phase current, the unipolar controller circuits can be used. Even if the drive is possibly operated with the non-sinusoidal current, DC current has the added advantage of reducing the hysteresis losses. The power converter plays a vital role in the industrial motor drive. The performance, size and the cost of the motor drive mainly depend upon the selected converter type of the power converter circuit.

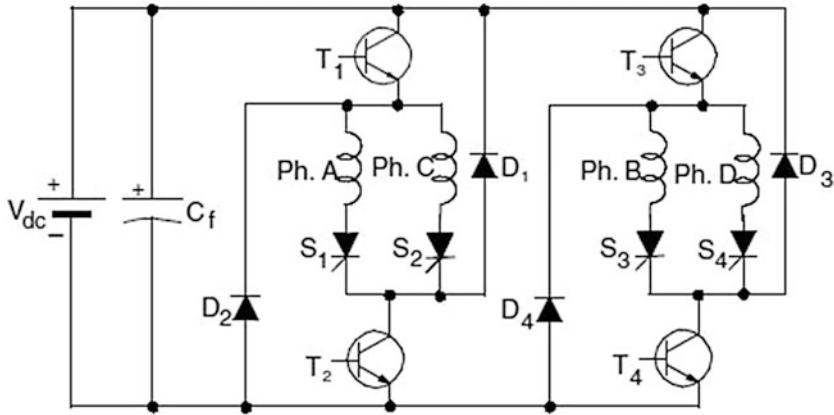


Fig. 7.2 Asymmetric power converter fed 6/4 SRM drive

7.3.1 Asymmetric Power Converter

With a specific end goal to have a fast generation of the excitation current, high switching voltage is required; Fig. 7.2 shows the asymmetric bridge converter considering only one phase of the SRM.

Figure 7.2 represents the 6/4 SRM model fed by the asymmetric bridge power converter. Here two power switches and two diodes per phase are connected to obtain the unipolar shift strategy.

7.3.2 R-Dump Power Converter

Figure 7.3 shows a converter formation with one diode and one transistor per each phase of the SRM.

7.3.3 C-Dump Power Converter

The C-dump converter fed SRM drive is shown in Fig. 7.3 with an energy recovery circuit. The residual magnetic energy in the machine phase has been transformed to mechanical energy (Fig. 7.4).

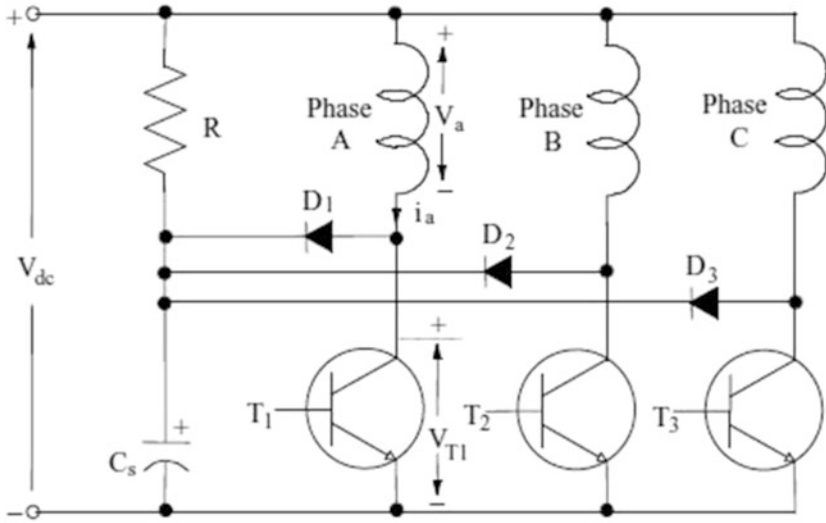


Fig. 7.3 R-dump power converter fed 6/4 SRM drive

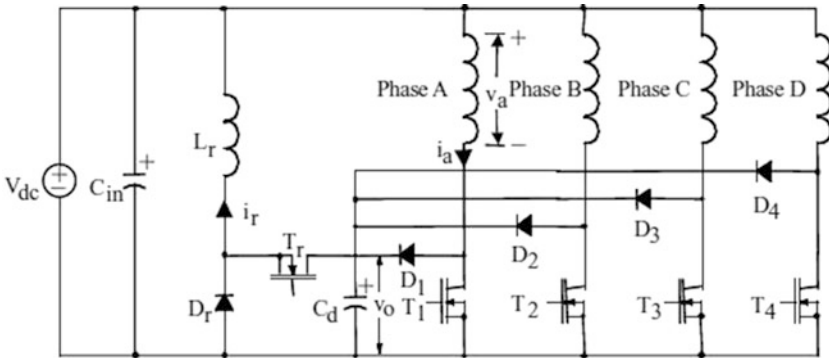


Fig. 7.4 C-dump converter fed 6/4 SRM drive

7.3.4 Fuzzy Logic Controller

In order to design the non-linear controllers, fuzzy logic was the appropriate choice because of its analytic dependence which are well justified by the universal approximation theorem [14]. Fuzzy control is one of the appropriate control schemes for torque control of SRM drives (Fig. 7.5).

Fig. 7.5 Fuzzy logic control scheme

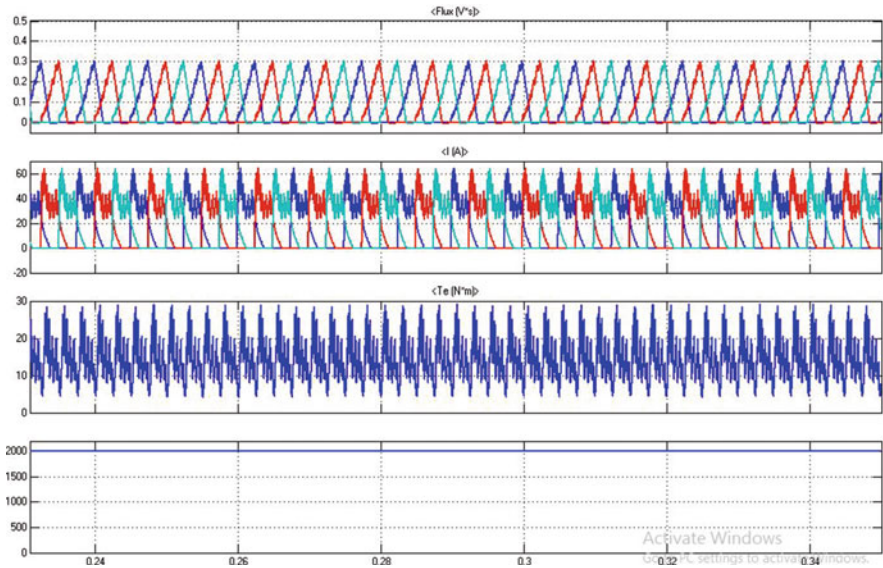
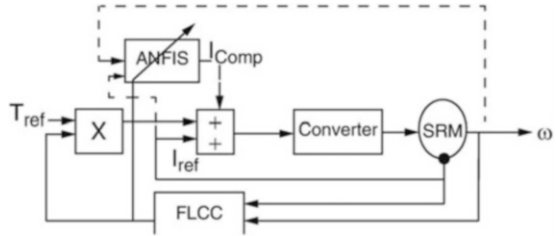


Fig. 7.6 Inductance profile, current, torque and speed of asymmetric converter fed SRM drive

7.4 Results and Discussions

Figure 7.6 shows the inductance profile, current, torque and speed of the SRM with different converters are discussed. Here the results are compared with Table 7.2 with its mean and standard deviation values. In this work, a three-phase 6/4 SRM motor is used in simulation by MATLAB/Simulink environment. The load torque is taken as 3 Nm and the speed of SRM is 2000 RPM. The main specifications of SRM are given in Table 7.1. The simulation is carried out for 0.35 s (Fig. 7.7).

Table 7.1 SRM simulation parameters

SRM parameters	Values
Stator poles	6
Rotor poles	4
Power	1 HP
Voltage	350
Current	70 A
Flux linkage	0.486
Stator resistance	0.05 ohm
Inertia	0.02 kg·m ²
Friction co-efficient	0.0183 N·m·s
Unaligned inductance	670 H
Aligned inductance	236 H
Saturated inductance	150 H
Maximum current value	450 A
Turn-on angle	25°
Turn-off angle	45°

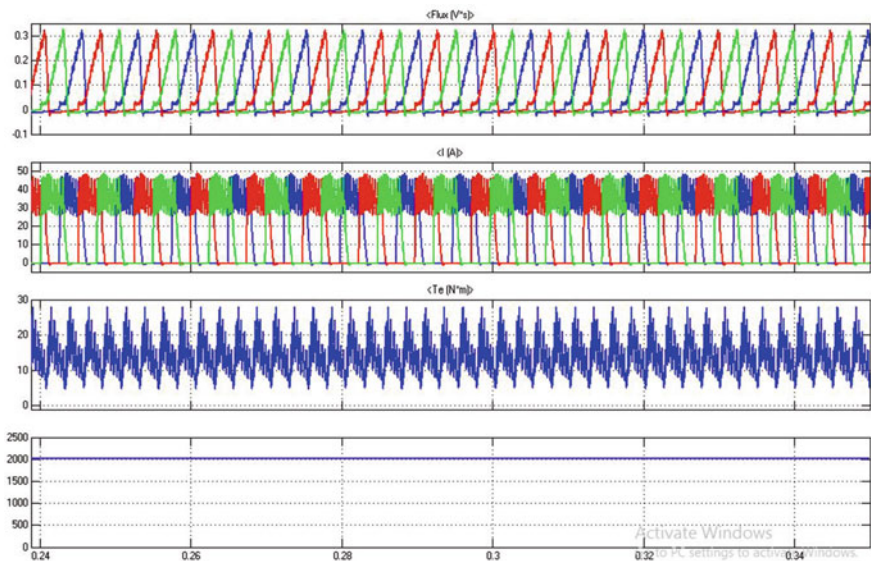


Fig. 7.7 Inductance profile, current, torque and speed of R-dump converter fed SRM drive

From the results shown, the ripples in the current waveforms are drastically reduced, which results in lesser torque ripples.

From the simulation results, the speed control can be achieved by fuzzy logic technique with minimum settling time and the parameters of SRM like speed and torque for the different converters are compared in Table 7.2 and (Fig. 7.8)

Table 7.2 Performance parameter comparison of various converter topologies for SRM drive

Power converter	Asymmetric power converter	R-dump converter	C-dump converter
Max value (T)	31.2856	27.9735	28.5908
Min value (T)	4.3225	4.9018	4.1015
Mean value (T)	14.0822	15.3658	16.1285
Standard deviation	4.3825	5.1134	5.8528
Torque ripple coefficient	0.5499	0.8247	0.8565
Setting time	0.014	0.020	0.028
Turn-on angle	45	45	45
Turn-off angle	75	75	75

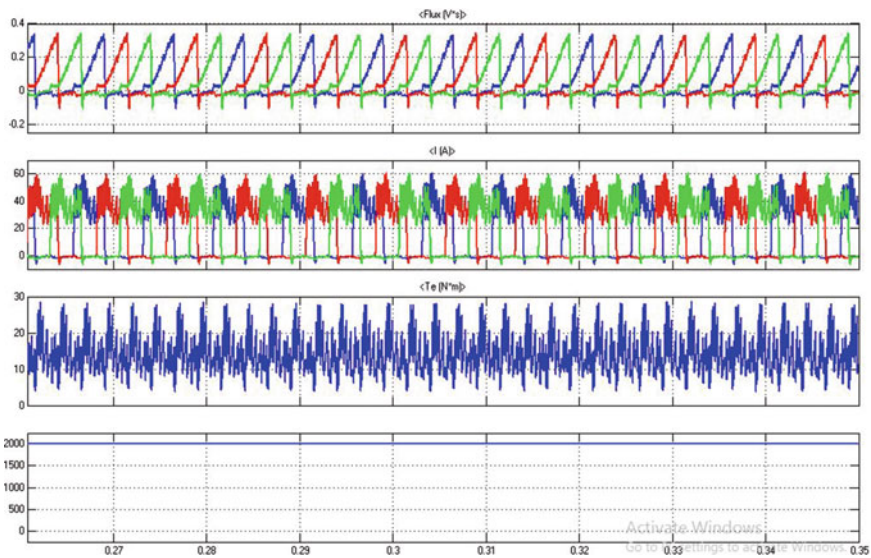


Fig. 7.8 Inductance profile, current, torque and speed of C-dump converter fed SRM drive

The statical performance obtained from the Table 7.2, such as minimum, maximum and mean values of total torque and torque ripple co-efficient for different types of power converters are listed, and these are compared. From the results, it is observed that mean torque is increased, and the torque ripple coefficient and standard deviation of total torque are reduced for asymmetric power converter fed SRM drive when compared with C-dump and R-dump converter fed SRM drive. Asymmetric power converter gives the robust performance, and it is the appropriate choice for the high-voltage applications as well as servo applications. The speed response is also appreciable for the asymmetric power converter over another converter, which results that this will be useful for the high-speed applications.

7.5 Conclusion

This paper discussed the different types of the power converters fed switched reluctance motor for minimizing its torque ripple along with the speed control. The various converters fed switched reluctance motor were developed in MATLAB Simulink environment such as asymmetric, R-dump and C-dump. Each converter current modulation technique has been done on a closed-loop control mechanism using a PID controller. The PID controller was tuned by the intelligent control techniques. The primary setback of SRM is higher torque ripple which was minimized by using current control technique with intelligent methods. The comparison table was constituted with mean, standard deviation, torque ripple coefficient and speed settling time values. From that, the asymmetric converter has relatively good performance over another converter for high-voltage and high-speed applications.

References

1. Krishnan R (2001) Switched reluctance motors drives, (modelling, simulation, analysis, design and application). CRC Press LLC, Boca Raton
2. Ma C, Qu L (2015) Multi objective optimization of switched reluctance motors based on design of experiments and particle swarm optimization. *IEEE Trans Energy Convers* 30(3):1144–1153
3. Kalaivani L, Subburaj P, Willjuice Iruthayarajan M (2014) Artificial intelligence-based control for torque ripple minimization in switched reluctance motor drives. *Acta Scientiarum Technol* 36:33–40
4. Ye J, Bilgin B, Emadi A (2015) An offline torque sharing function for torque ripple reduction in switched reluctance motor drives. *IEEE Trans Energy Convers* 30(2):726–735
5. Cao G-Z, Fang JL, Huang S-D (2014) Optimization design of the planar switched reluctance motor on electromagnetic force ripple minimization. *IEEE Trans Magn* 50(11):1–4
6. Mikil R, Husain I, Islam MS, Sozer Y (2015) Four-quadrant torque ripple minimization of switched reluctance machine through current profiling with mitigation of rotor eccentricity problem and sensor errors. *IEEE Trans Ind Appl* 51(3):2097–2104
7. Ye J, Bilgin B, Emadi A (2015) An extended-speed low-ripple torque control of switched reluctance motor drives. *IEEE Trans Power Electron* 30(3):1457–1470
8. Kalaivani L, Subburaj P, Willjuice Iruthayarajan M (2013) Speed control of switched reluctance motor with torque ripple reduction using non-dominated sorting genetic algorithm (NSGA-II). *Int J Electr Power Energy Syst* 53:69–77
9. Vujić VP (2012) Minimization of torque ripple and copper losses in switched reluctance drive. *IEEE Trans Power Electron* 27(1):388–399
10. Laboid C, Srairi K, Mahdad B (2015) Speed control of 8/6 switched reluctance motor with torque ripple reduction taking into account magnetic saturation effects. *Energy Proc* 74:112–121
11. Suryadevara R, Fernandes BG (2013) Control techniques for torque ripple minimization in switched reluctance motor: an overview. *ICIIS*, pp 18–22

12. Sozer Y, Husain I, Torrey DA (2015) Guidance in selecting advanced control techniques for switched reluctance machine drives in emerging applications. *IEEE Trans Ind Appl* 51 (6):4505–4514
13. Huang C, Shao Z, Zhao Y (2003) Application of neural network to suppress PD stochastic pulse noise for large generators. *IEEE Trans*, pp 21–25
14. Ali Akcayol M (2004) Application of adaptive neuro-fuzzy controller for SRM. *Adv Eng Softw* 5:129–137

Chapter 8

Design of Half-Ring MIMO Antenna to Reduce the Mutual Coupling



K. Vasu Babu and B. Anuradha

Abstract We propose a novel, dual-polarized MIMO radiator, which consists of half rings, and included in the ring a square shape at the top side and half circular shape at bottom side operating from 2 to 10 GHz with a microstrip feeding. The suppression of mutual coupling is obtained by maintaining the separation between the patches around $0.25 \lambda_0$. The half-ring MIMO antenna is resonating at dual band of frequencies at 7.24 and 8.16 GHz with impedance bandwidths of 430 and 230 MHz, has $|S_{11}| < -10$ dB in the MIMO range from 2.0 to 10.0 GHz and has a mutual coupling with $|S_{21}| < -20$ dB. The radiator has a low ECC (envelope correlation coefficient) with values equal to approximately less than 0.025, which will prove that the half-ring MIMO radiator shows better diversity performance. The half-ring MIMO radiator has improved the parameters of reflection coefficient, mutual coupling, realized gain, group delay and real/imaginary impedances.

Keywords Multiple input-multiple output · Realized gain · ECC · Group delay · Real/imaginary impedances

Abbreviation

ECC	envelope correlation coefficient
GD	group delay
MIMO	multiple input-multiple output
NGD	negative group delay
UWB	ultra-wide band antenna

K. V. Babu (✉)
Department of ECE, VVIT, Guntur, Andhra Pradesh, India

B. Anuradha
Department of ECE, Sri Venkateswara University, Tirupati, Andhra Pradesh, India

8.1 Introduction

The major challenge in the design of any multiple input-multiple output (MIMO) antenna is when the antenna elements are at close proximity, it produces a mutual coupling. A few isolation techniques like using split ring resonators reduced the isolation to 20 dB [1], and by using common mode suppression of ultra-wide band antenna (UWB)-MIMO radiator reduces the isolation to 15.5 dB [2] and maintains an ECC value of 0.04. To reduce the mutual coupling, the design structure Minkowski produced multi-band [3] characteristics having S_{21} of 42 dB. A compact antenna having an area $23 \times 39.5 \text{ mm}^2$ and polarization diversity array produced an isolation of 20 dB operated between 2.5 and 12 GHz [4] whose ECC value maintained at less than 0.55. In [5] a parasitic approach like a UWB-MIMO radiator used a carbon film black operated from 2.5 to 11 GHz with reflection coefficient (S_{21}) of 20 dB. A floating parasitic decoupling structure [6] with a size of 33×45.5 produced a UWB system whose isolation ≤ 20 dB is achieved. An extendable MIMO array antenna has a low ECC around 0.025 and its peak gain 4 dBi was observed [7]. A wide-band MIMO antenna with band rejection [8], tri-band MIMO for defence and radio astronomy applications [9] was observed. For the mobile terminal applications, a compact G-shaped design [10] having an area of $50 \times 82 \text{ mm}^2$, operating in the frequency ranges from 2.2 to 13.3 GHz with isolation of ≤ 15 dB and a maximum gain of 4.78 dBi is achieved. Finally, it is interesting to observe that the proposed half-ring design structure reduced the mutual coupling to 40 dB. By using the techniques of optimization procedure finalized the dimensions. Table 8.1 shows the comparison with previous designs.

8.2 Antenna Construction and Its Geometry

The proposed structure having dimensions $46 \times 35 \text{ mm}^2$ is designed with a thickness of 1.5 mm on FR-4 substrate. The loss tangent ($\tan \delta$) is 0.02 and relative permittivity (ϵ_r) of the substrate is 4.3. The half-ring MIMO system produces a dual-band

Table 8.1 Comparison with existing systems

Refs.	Size (mm ²)	S_{11} (dB)	S_{21} (dB)	Gain (dBi)	f_L - f_H (GHz)
Irene and Rajesh [1]	40.5×40.5	45	20	6	3.1-10.6
Liu and Tu [2]	44×44	39	15.5	3.8	2.9-10.8
Braaten et al. [4]	23×39.5	27	20	3.6	2.0-10.0
Chen et al. [5]	50×40	32	15	4.73	2.5-11
Koul et al. [7]	38×90	20	20	4	3.1-16
Kim et al. [8]	55×13.5	35	17.2	4.2	3.1-10.6
Toktas [10]	50×82	34	15	4.78	2.2-13.3
Proposed system	46×35	29.25	21.89	4.75	2.0-10.0
		12.85	40	5	2.0-10.0

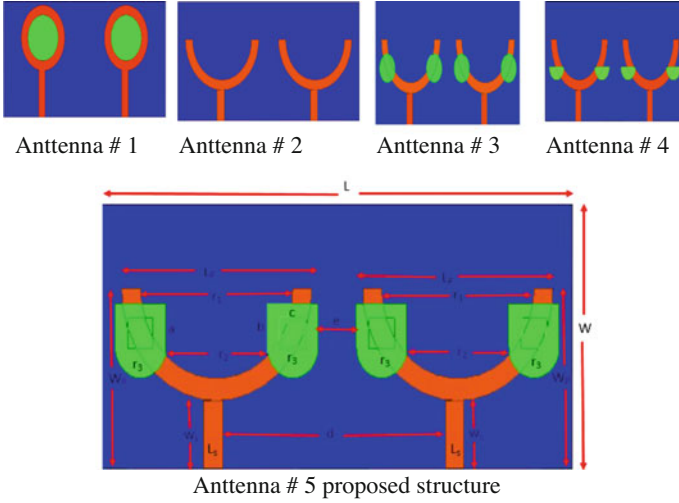


Fig. 8.1 Evolution stages of half-ring antenna

Table 8.2 Dimensions of half-ring MIMO antenna (in mm)

Parameter	L	W	L_p	W_p	r_1	r_2	r_3
Dimensions (mm)	50.5	35	38	25	3	10	2
Parameter	A	b	c	d	E	W_s	L_s
Dimensions (mm)	10	9	7	7	4	18	10

structure using the C-band applications. The step-by-step process of analysis is indicated in Fig. 8.1. Antenna # 1 shows that a strip is added to a circular ring which produced a reflection coefficient of 22 dB. Antenna # 2 represents the cutting of the half of the circles on both sides of total ring with S_{11} 13 dB. Antenna #3 represents adding half semi-circles on both sides of the patch and S_{11} is reached to 34 dB. Antenna #4 having a square patch is added to half circles on both sides to form a new structure whose S_{11} is 10 dB. Finally, Antenna #5 represents the proposed structure having the reflection coefficient greater than 25 dB and transmission coefficient of 40 dB. The design dimensions of the half-ring MIMO structure are listed in Table 8.2.

Finally, the designed structure having the dual band of operation produces the parameters S_{11} and S_{21} maintained less than 10 and 20 dB. In this design, strengthening the parameter mutual coupling between the two symmetrical radiators distance can be considered as 4 mm.

Figure 8.2 indicates the S -parameter evolution of each stage of half-ring MIMO design which shows all the cases $S_{11} \leq -10$ dB. Figure 8.3 shows the comparison of S parameters (S_{11} and S_{21}), which shows that a greater improvement in the mutual coupling observed at 7.2 GHz is 21.89 dB and at 8.1 GHz it almost touches to 40 dB reduction. Figure 8.4 shows the S_{21} comparison at different distances. This result shows that a greater interference between the patches is reduced. Figure 8.5 depicted

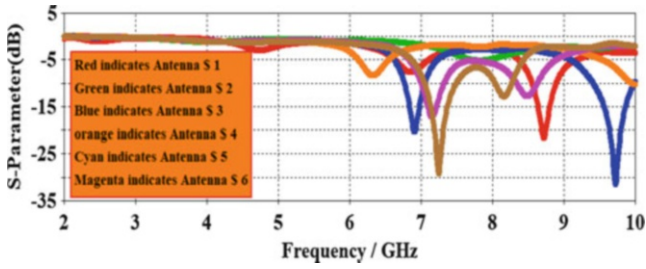


Fig. 8.2 Step-by-step analysis of S parameters

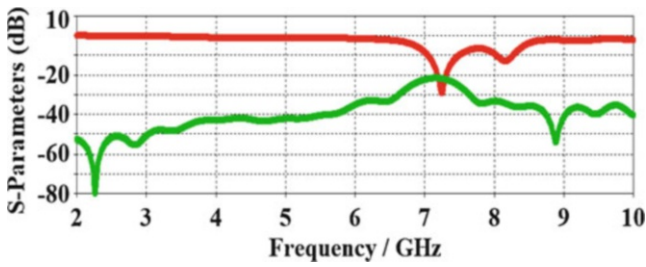


Fig. 8.3 S parameter comparison

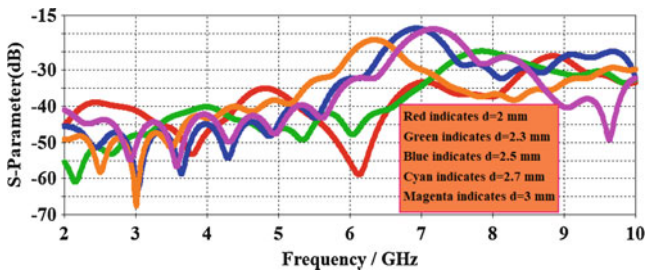
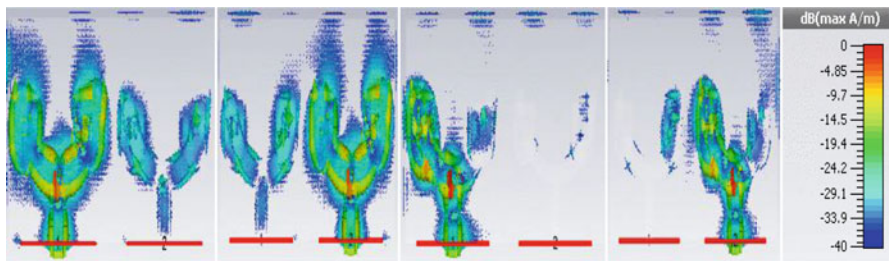


Fig. 8.4 S_{21} Graph at different spacing between antennas



a) Port 1 excited 7.2 GHz b) Port 2 excited 7.2 GHz c) Port 1 excited 8.1 GHz d) Port 2 excited 8.1 GHz

Fig. 8.5 Surface current distribution of the design

the distribution of surface current vectors at 7.2 GHz as well as at 8.1 GHz. Figure 8.5a perceived that port #1 is excited and port #2 is terminated with a characteristic impedance of 50Ω load. In this case we observed that the maximum current flows through half-ring structure and observed that within a half ring a semi-circle is added with a square patch maximum current flowing. Figure 8.5b perceived that port #2 is excited and port #1 is terminated with 50Ω load observed that half-ring MIMO structure at bottom of the ring structure patch maximum current is flowing. For both the ports, the analogous distribution of surface current perceived that two homogenous elements were positioned side by side which can lead to create first resonant mode at 7.2 GHz. Similarly, the second resonant mode is at 8.1 GHz. Figure 8.5c shows the distribution of surface current vectors, maximum current flows at the left side of the half-ring resonator and a partial current flows through the edge of the microstrip connector and a little amount of current is observed at the left side of the half-ring patch. Figure 8.5d identified that maximum current flows through the left-side patch, a partial current is observed at the edge of the strip and a minimum surface current on right-side patch having a half-ring added to the square patch.

8.3 Results Analysis and Discussion

The simulated representation of S parameters satisfies the condition at dual band of operations of $|S_{11}| < -10$ dB and $|S_{11}| < -20$ dB. These parameters indicate that the proposed structure has a greater improvement in the parameters of transmission coefficient as well as reflection coefficients at the dual-band resonant frequencies of 7.2 and 8.1 GHz for the applications of C-band and X-band. Figure 8.6 shows the real/imaginary part of impedance verses frequency from 2.0 to 10 GHz. Figure 8.7 indicates the co- and cross-polarization effects with respect to the xz and yz planes having a dual-band operation at 7.2 and 8.1 GHz when one port of the antenna is excited and another port of the system is terminated with characteristic impedance of

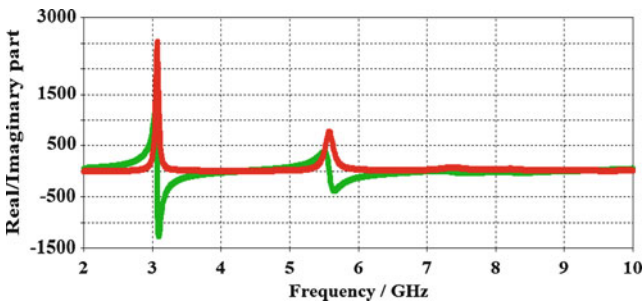


Fig. 8.6 Real/imaginary impedances

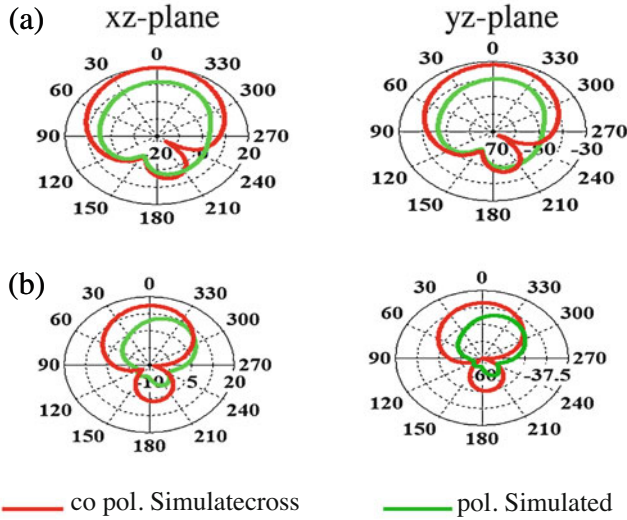


Fig. 8.7 Radiation patterns at dual-band (a) $f_1 = 7.2$ GHz and (b) $f_2 = 8.1$ GHz

50 Ω . It is observed that the pattern of radiation in xz plane is approximately analogous to pattern in the yz plane. It is important to identify that co-pol patterns are omnidirectional in nature in xz plane. Similarly, it is also to identify that the cross-pol component magnitude increases as the response of the frequency increases with horizontal direction component. Here, we observed that the co-pol is always outer region of the radiation pattern and cross-pol is inner side of the co-pol is a reasonably good agreement. Therefore, it can say that in order to placing half-ring antennas placing linearly the radiation patterns of separate elements lies on surface of a plane similar to each other by properly choosing the separation between the two radiating patches. It is also realized that the proposed structure helps to reduce the factor of mutual coupling.

Figure 8.8 shows the realized gain variations of the half-ring structure produced dual-band operation when port #1 is excited and port #2 is terminated with a 50 Ω load. When port #1 is excited, peak gain in first resonant band is 6 dBi and at another band is 4.8 dBi. Figure 8.9 shows the parameter of ECC which is lower than the 0.03. In time-domain analysis process group delay is a major important parameter. It shows how much time is considered for antenna to receive a pulse. There is a variation in group delay when placing the antenna front to front and side by side for 1 n. It validates the linear phase over entire far-field region. Figure 8.10 indicates the simulated result of group delay of 7.85 ± 1 n, -0.98 ± 1 ns over a resonant band of frequencies at 7.2 and 8.1 GHz. From the graph it is interesting to observe that the factor of negative group delay (NGD) increases, the signal attenuation increases and the bandwidth decreases.

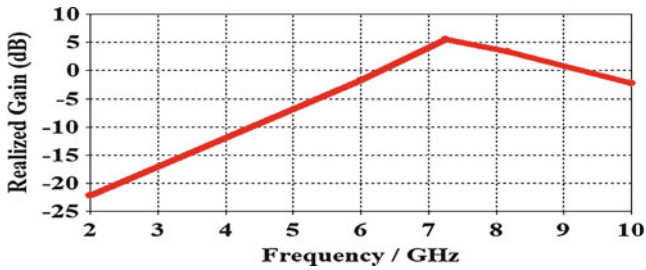


Fig. 8.8 Comparison of realized gain

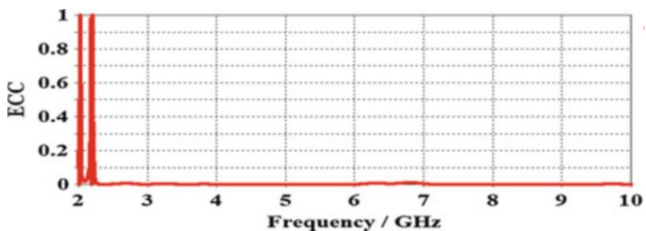


Fig. 8.9 ECC Simulation at 2.5 and 6.4 GHz

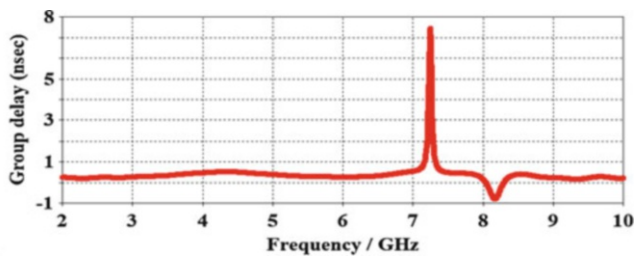


Fig. 8.10 Group delay of proposed design

8.4 Conclusion

A half-ring MIMO antenna, which consists of semi-half-ring added to a square patch, has been produced with dual-band characteristics at 7.2 and 8.1 GHz for C-band and X-band applications, respectively. The ineluctable suppression of the isolation has been attained by selecting the distance between the patches nearly $0.25 \lambda_0$. We observed that from the simulation results for the dual-band region $|S_{11}| < -10$ dB has been achieved and impedance matching maintained at $|S_{21}| < -20$ dB in the dual resonant frequencies. The group delays (GD) at the dual-band frequencies are 7.85 ± 1 ns and -0.98 ± 1 ns, and ECC value lower than

the 0.03 has been identified for the half-ring design. Thus, this system arrangement improves the parameter of antenna and good radiation patterns have been observed.

References

1. Irene G, Rajesh A (2018) A dual-polarized UWB–MIMO antenna with IEEE 802.11ac band-notched characteristics using split-ring resonator. *J Comp Electron*. <https://doi.org/10.1007/s10825-018-1213-x>
2. Liu YY, Tu ZH (2017) Compact differential band-notched UWB-MIMO antenna with common-mode suppression. *IEEE Antennas Wirel Propag Lett* 16:593–596
3. Babu KV, Anuradha B (2018) Design of multi-band Minkowski MIMO antenna to reduce the mutual coupling. *J King Saud Univ Eng Sci* 35:1–7
4. Braaten BD, Khan MS, Ijaz B, Capobianco AD, Naqvi A, Asif S (2015) Planar, compact ultra-wideband polarization diversity antenna array. *IET Microw Antennas Propag* 9:1761–1768
5. Chen LS, Lin GS, Sung H, Houng MP (2017) Isolation improvement in UWB MIMO antenna system using carbon black film. *IEEE Antennas Wirel Propag Lett* 16:222–225
6. Najam AI, Khan MS, Capobianco AD, Shafique MF, Shoaib I, Autizi E (2014) Compact ultra-wideband diversity antenna with a floating parasitic digitated decoupling structure. *IET Microw Antennas Propag* 8:747–753
7. Koul SK, Sipal D, Abegaonkar MP (2017) Easily extendable compact planar UWB MIMO antenna array. *IEEE Antennas Wirel Propag Lett* 16:2328–2331
8. Ryu KB, Lee JM, Kim HK, Woo JM (2012) A compact ultra wideband MIMO antenna with WLAN band-rejected operation for mobile devices. *IEEE Antennas Wirel Propag Lett* 11:990–993
9. Babu KV, Anuradha B (2018) Tri-band MIMO antenna for WLAN, WiMAX and defence system & Radio astronomy applications. *Adv Electromagn* 7:60–67
10. Toktas A (2016) G-shaped band-notched ultra-wideband MIMO antenna system for mobile terminals. *IET Microw Antennas Propag* 11(5):718–725

Chapter 9

Optimal Allocation of Distributed Generation Using Clustered Firefly Algorithm



K. Banumalar, B. V. Manikandan, and S. Sundara Mahalingam

Abstract Integration of distributed generation units (DGs) in distribution systems aims to enhance the system performance. Location and the sizing are the two important factors on the network power loss. This work proposes a clustered firefly algorithm (CFFA) to reduce the distribution system loss, simultaneous optimal placement and sizing of the distributed generation resources in radial distribution systems studied. The simulation is done on IEEE-69 bus network in MATLAB software. The simulated results demonstrate the effectiveness of the proposed clustered firefly algorithm compared with other optimization algorithms.

Keywords Distributed generation (DG) · Optimal DG location · Optimal DG size · Loss minimization · Radial distribution system

Abbreviations

DG	distributed generation
CFFA	clustered firefly algorithm
PLI	power loss index
TPL_{DG}	total real power losses with DG
TPL	total real power losses without DG
P_i and P_j	net real power injection in bus 'i' and 'j'.
Q_i and Q_j	net reactive power injection in bus 'i' and 'j'.
R_{ij}	resistance between bus 'i' and 'j'.
V_i and V_j	voltage at bus 'i' and 'j'
δ_i and δ_j	angle at bus 'i' and 'j'
$ V_i ^{min}$ and $ V_i ^{max}$	minimum and maximum limit of voltage in bus 'i'.
$ I_{ij} ^{max}$	maximum limit of current in bus 'i' and 'j'.

K. Banumalar (✉) · B. V. Manikandan · S. Sundara Mahalingam
Department of Electrical and Electronics Engineering, Mepco Schlenk Engineering College,
Sivakasi, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_9

9.1 Introduction

In recent years, electric distribution systems are becoming large and complex leading to higher system losses and poor voltage regulation. Studies indicate that almost 10–13% of the total power generated is lost as I^2R losses at the distribution level. It is evident that any loss reduction is beneficial to distribution utilities, which is generally the entity responsible to keep losses at low levels. Loss reduction is, therefore, the most important factor to be considered in the planning and operation of distributed generation (DG) [1, 2].

In a radial feeder, depending on the technology, DG units can deliver a portion of the total real and/or reactive power to loads so that the feeder current reduces from the source to the location of DG units. However, studies [3] have indicated that if DG units are improperly allocated and sized, the reverse power flow from larger DG units can lead to higher system losses. Hence, to minimize losses, it is important to find the best location and size given the option of resource availability. The methodology applies the clustered firefly algorithm (CFFA) in order to minimize the system loss. Minimum system losses are obtained subjected to voltage constraint and current limit.

The organization of this paper is as follows: Section 9.2 addresses the problem formulation. The clustered firefly algorithm (CFFA) is implemented in optimal placement and sizing of the distributed generation represented in Sect. 9.3. Simulation results on the test systems are illustrated in Sect. 9.4. Section 9.5 presents the performance comparison. Then, the conclusion is given in Sect. 9.6.

9.2 Problem Formulation

Proposed methodology presented in this paper is aimed to optimize technique functionality of distribution system by minimizing the power losses in distributed generation units [4] given radial distribution network.

9.2.1 The Objective Function

Power loss index (PLI) is formulated by Eq. (9.1) as.

Minimize

$$PLI = \frac{TPL_{DG}}{TPL_{DG} - TPL} \quad (9.1)$$

where TPL_{DG} and TPL are the total real power losses with and without DG units in the distribution system, respectively. The total real power loss (TPL) is calculated using Eq. (9.2):

$$TPL = \sum_{i=1}^n \sum_{j=1}^n A_{ij}(P_i P_j + Q_i Q_j) + B_{ij}(Q_i P_j - P_i Q_j), \quad (9.2)$$

where.

$$A_{ij} = \frac{R_{ij} \cos(\delta_i - \delta_j)}{V_i V_j}. \quad (9.3)$$

$$B_{ij} = \frac{R_{ij} \sin(\delta_i - \delta_j)}{V_i V_j}. \quad (9.4)$$

where P_i and P_j are the net real power injection in bus 'i' and 'j', respectively. Q_i and Q_j are the net reactive power injection in bus 'i' and 'j', respectively. R_{ij} is the line resistance between bus 'i' and 'j'. V_i and V_j are the voltage at bus 'i' and 'j', respectively. δ_i and δ_j are the angle at bus 'i' and 'j', respectively.

The above problem formulation is solved subject to the constraints listed below.

9.2.2 Voltage Limits

The voltage must be kept within standard limits at each bus.

$$|V_i|^{\min} \leq |V_i| \leq |V_i|^{\max}. \quad (9.5)$$

9.2.3 Current Limits

$$|I_{ij}| \leq |I_{ij}|^{\max} \quad (9.6)$$

where $|V_i|^{\min}$ and $|V_i|^{\max}$ are the minimum and maximum limit of voltage in bus 'i'. $|I_{ij}|^{\max}$ is the maximum limit of current in bus 'i' and 'j'.

9.3 Application of CFFA for Power Loss Minimization

The most effective algorithms for optimization problems are nature-inspired algorithm. Firefly algorithm is a new swarm-based algorithm and it is proposed based on the behaviour of fireflies. It is developed by Yang [5]. This firefly algorithm is implemented in clustered based environment to maintain the balance between exploration and exploitation and it is referred as clustered firefly algorithm (CFFA) [6].

Figure 9.1 shows the flowchart for solving placement and sizing problem using CFFA. The implementation steps are given for CFFA:

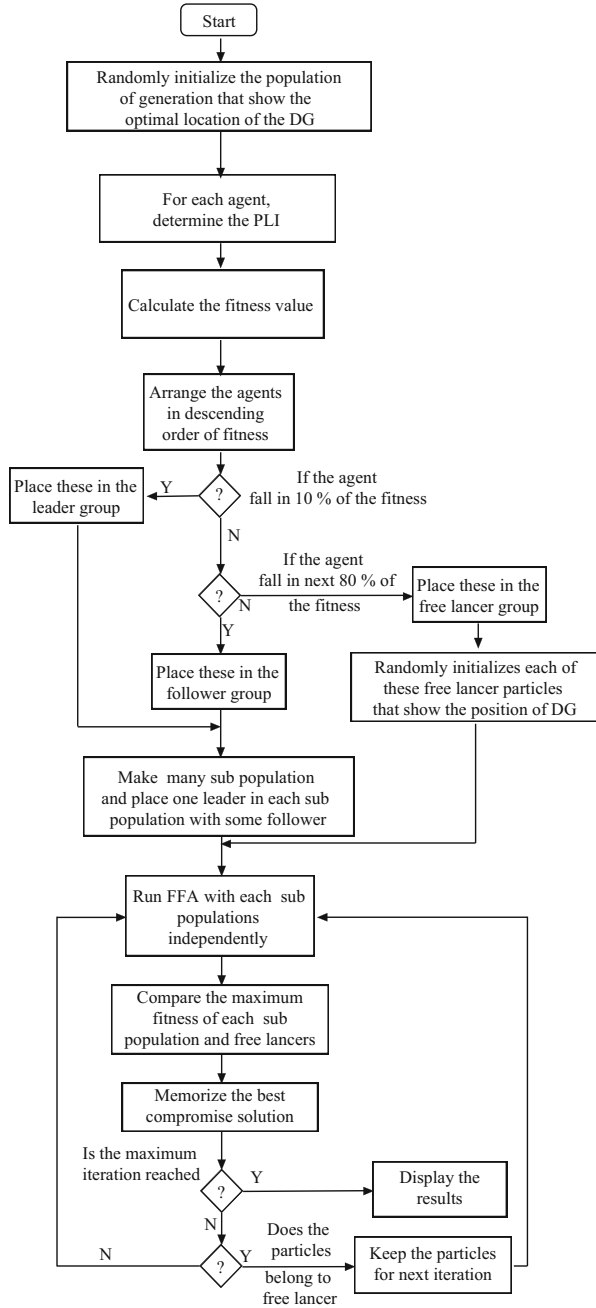
1. Reading the total information of the network including the number of buses, line flow, real and reactive power loss of the distribution system without any DG units etc.
2. Identifying the adjustable parameters of CFFA and the maximum iteration number.
3. Generating the initial population and initializing the firefly positions in the search space.
4. Computing the performance indices such as PLI.
5. Determining the fitness function of each firefly. Memorize the best solution based on maximum fitness value of the firefly.
6. Updating the position of each firefly.
7. Repeating the steps 4–6 until reaching the termination criteria (maximum number of iteration).
8. Reporting the best value of the memory.

9.4 Simulation Results and Discussion

To study the functionality of CFFA in the problem of simultaneously optimal location and sizing of distributed generation resources in radial distribution systems with the goal of minimizing the loss and the simulation is conducted on the IEEE 69-bus systems (Fig. 9.2) in MATLAB software. The optimal parameters used to solve optimal sizing and placement problem are given in Table 9.1.

Tables 9.2 and 9.3 present the results of optimal sizes and locations of single DG and multi-DG units by CFFA. In order to illustrate the performance of the proposed method, the performance of CFFA is compared with other techniques available in the literature and is presented in Tables 9.2 and 9.3.

Fig. 9.1 Flowchart for DG placement and sizing problem using CFFA



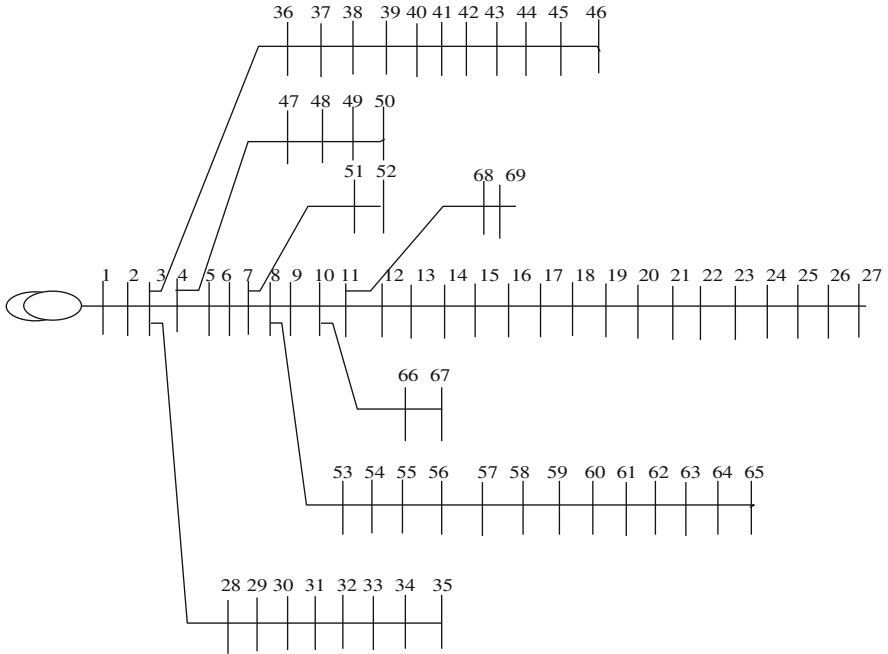


Fig. 9.2 Single line diagram of IEEE standard 69-bus radial distribution networks

Table 9.1 Control parameters of 69-bus system for optimal sizing and placement of DG

CFFA parameters	Parameter value
Population size	100
No. of iterations	300
γ	1.5
β_0	0.5
Constant, n	1
No. of clusters	8
No. of leaders	10
No. of followers	80
No. of free lancers	10

From Tables 9.2 and 9.3, it is perceived that the performance of the CFFA is better in terms of power loss minimization.

Table 9.2 Placement and sizing of single DG

Methods	Objective function- power loss (kW)	DG location and size (kW)
PSI [7]	83.142	61(1863.10)
Novel power loss sensitivity [7]	83.195	61(1832.45)
VSI search techniques [7]	83.139	61(1870.00)
Dynamic programming method [8]	83.337	61(1829.78)
Crow search algorithm [9]	83.224	61(1873)
CFFA	83.012	61(1872.26)

Table 9.3 Placement and sizing of multi-DG

Methods	Power loss (kW)	Multi-DG location and size (KW)
Genetic algorithm [10]	89.0	21(929.7), 62(1075.2), 64(984.8)
Dynamic programming method [8]	88.21	60(753.89), 31(800.1), 15(1294.6)
Particle swarm optimization [10]	83.2	61(1199.8), 63(795.6), 17(992.5)
GA-PSO [10]	81.1	63(884.9), 61(1192.6), 21(910.5)
Water cycle algorithm [11]	71.5	61(775), 62(1105), 23(438)
Crow search algorithm [9]	69.429	11(535.1), 18(377), 61(1717.5)
CFFA	69.12	11(527.5), 18(380.0), 61(1719.2)

9.5 Performance Comparison

From Fig. 9.3, we observed that the CFFA have some advantages over GA (genetic algorithm), PSO (particle swarm optimization) and DP method. By comparing with above algorithm, the advantages of CFFA are given as follows:

- The population in CFFA can be automatically divided into subgroups so that these subgroups search around the multimodal optima. This makes it possible for the algorithm to find all global optima simultaneously as in Fig. 9.3. Thus, the algorithm is particularly suitable for nonlinear, non-smooth multimodal optimization problems.
- The number of control parameters that need initial tuning is less than GA and PSO. Also, a suitable selection scheme has to be adopted in GA.
- The probability of obtaining feasible solution is high.

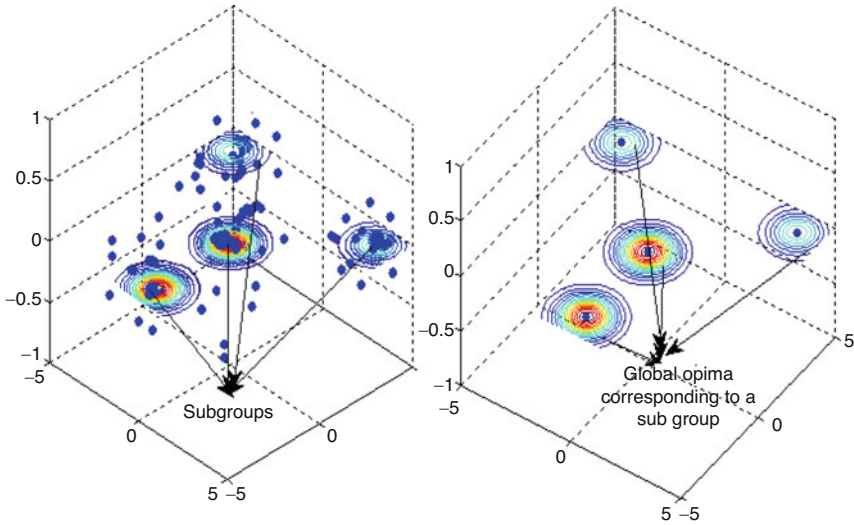


Fig. 9.3 Convergence behaviour of CFFA

9.6 Conclusion

This paper has presented CFFA method for single and multiple DG allocation for loss reduction. The methodology is tested on a 69-bus system. However, the proposed method is very efficient in finding the optimal solution. The results are compared with the results of GA, PSO, DP, WCA (water cycle algorithm) and CSA (common scrambling algorithm) available in the literature. The computational results showed that the performance of the CFFA is better than other algorithm available in the literature. The proposed method can be easily applied and adapted to any large-scale radial distribution networks.

References

1. Ng H, Salama M, Chikhani A (2000) Classification of capacitor allocation technique. *IEEE Trans Power Deliv* 15(1):387–392
2. Singh D, Verma KS (2009) Multiobjective optimization for DG planning with load models. *IEEE Trans Power Syst* 24(1):427–436
3. Atwa YM, El-Saadany EF, Salama MMA, Seethapathy R (2010) Optimal renewable resources mix for distribution system energy loss minimization. *IEEE Trans Power Syst* 25(1):360–370
4. Nikmehr N, Najafi-Ravadanegh S (2015) Optimal power dispatch of multi-microgrids at future smart distribution grids. *IEEE Trans Smart Grid* 6(4):1648–1657
5. Yang XS (2009) Firefly algorithms for multimodal optimization. *Lect Notes Comput Sci* 5792:169–178

6. Banumalar K, Manikandan BV, Arul Jeyaraj K, Chandrasekaran K (2018) Optimal placement of phasor measurement units using clustered gravitational search algorithm. *J Intell Fuzzy Syst* 34:4315–4330
7. Murty VVSN, Kumar A (2015) Optimal placement of DG in radial distribution systems based on new voltage stability index under load growth. *Electr Power Energy Syst* 69:246–256
8. Khalesi N, Rezaei N, Haghifam MR (2011) DG allocation with application of dynamic programming for loss reduction and reliability improvement. *Electr Power Energy Syst* 33:288–295
9. Barati H, Shahsavari M (2018) Simultaneous optimal placement and sizing of distributed generation resources and shunt capacitors in radial distribution systems using crow search algorithm. *Int J Ind Electron Control Optim* 1(1):27–40
10. Moradi MH, Abedini MA (2012) Combination of genetic algorithm and particle swarm optimization for optimal DG location and sizing in distribution systems. *Electr Power Energy Syst* 34:66–74
11. Adel A, El-Ela A, Ragab A, El-Sehiemy ASA (2018) Optimal placement and sizing of distributed generation and capacitor banks in distribution systems using water cycle algorithm. *IEEE Syst J* 22(4):1–8

Chapter 10

CDM-Based Two-Degree-of-Freedom PID Controller Tuning Rules for Unstable FOPTD Processes



Somasundaram S and Benjanarasuth T

Abstract This paper deals with the coefficient diagram method (CDM)-based two-degree-of-freedom proportional integral derivative (CDM-PID) controller tuning rules for unstable first-order plus time delay (UFOPTD) processes. The explicit tuning rules for setting the PID controller parameters are derived using a general UFOPTD transfer function model, the second-order Taylor denominator approximation technique and the pole allocation strategy named CDM. The derived tuning rules are novel, directly relating the controller parameters to the process model parameters. Simulation results indicate that the CDM-PID controller with the proposed tuning rules deliver better performance than other available PID controller tuning methods.

Keywords CDM · PID · 2DOF · UFOPTD · Tuning rules

Abbreviations

CDM	coefficient diagram method
PID	proportional integral derivative
2DOF	two-degree-of-freedom
UFOPTD	unstable first-order plus time delay
τ	equivalent time constant
γ_i	stability indices

Somasundaram S (✉)

Department of Electronics & Instrumentation Engineering, Annamalai University,
Chidambaram, Tamil Nadu, India
e-mail: somasundaram@accetedu.in

Benjanarasuth T

Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok,
Thailand
e-mail: taworn.be@kmitl.ac.th

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_10

st	settling time
K	process gain
T	process time constant
L	process time delay
K_c	proportional gain
T_i	integral time
T_d	derivative time
%OV	percentage overshoot
TV_{sum}	sum of the total variation

10.1 Introduction

Despite significant developments in control theory and technology, the classical PID controllers are still popular and remain the most widely [1] used controller in the process industries. This is largely due to their structural simplicity, reliability and satisfactory control over wide class of processes.

In industries, there exist many important chemical and biological processes [2] that exhibit multiple steady states [3] due to the nonlinear dynamic behaviour, uncertainty and time-varying parameters. To maximize product quality and to maintain safety, it is often desirable to operate such processes within certain unstable steady state [4] boundaries. The open loop dynamics of such processes can be adequately described by unstable first-order plus time delay (UFOPTD) transfer function model [4].

Readers may refer to [2, 5–7 and the references therein] addressing some of the recent methods of designing PID controllers for the processes modelled as UFOPTD. Although there exist PID controller design methods based on UFOPTD models, there is still a strong interest in the design and tuning of a new, simple and efficient PID controller for UFOPTD processes.

The objective of the present research is that, given a UFOPTD transfer function model, a simple and effective PID controller must be designed in such a way that the control system achieves a good time response with satisfactory performance. To realize this, the concept of coefficient diagram method (CDM) [8] is used.

Even though there exist CDM-based PID [9, 10] controllers, it is emphasized that the design and tuning of CDM-PID controller presented in this work is different and new to the literature. It should be noted that the existing design methods presented in refs. [9, 10] involve complex algebra and lengthy computational steps. Hence, the controller design is hard, tedious and time consuming. In this work, the problem of designing CDM-PID controller is addressed by means of simple and novel tuning rules. The advantage and key feature of the proposed tuning rules are that they offer more flexibility for the designer to compute the controller parameters and also make the CDM-PID design easier when compared to the existing methods.

10.2 Coefficient Diagram Method

Coefficient diagram method (CDM) is a polynomial design method. The simultaneous approach of CDM helps the designer to maintain a good balance between the coefficients of the controller polynomials and the design specifications. For more information on CDM, readers can refer to [11]. The CDM control structure is shown in Fig. 10.1.

In Fig. 10.1, r , e , u , d and y represent the reference input, error value, controller output, output disturbance and the process output, respectively. $N(s)$ and $D(s)$ are the numerator and denominator polynomials of the transfer function of the process which is to be controlled. $A(s)$, $F(s)$ and $B(s)$ are the forward denominator polynomial, set point filter and the feedback numerator polynomials of the CDM controller, respectively. The equation relating the input (r) and the disturbance (d) to the process output (y) is written as

$$y = \frac{N(s)F(s)}{P(s)}r + \frac{A(s)D(s)}{P(s)}d \tag{10.1}$$

Here, $P(s)$ is the closed-loop characteristic polynomial and is defined by

$$P(s) = A(s)D(s) + B(s)N(s) \tag{10.2}$$

The controller polynomials $A(s)$ and $B(s)$ are given as.

$$A(s) = \sum_{i=0}^p l_i s^i \text{ and } B(s) = \sum_{i=0}^q k_i s^i; \quad p \geq q \tag{10.3}$$

where l_i and k_i are the coefficients of the controller polynomials.

Upon substituting (10.3) in (10.2), the polynomial $P(s)$ is expressed as follows:

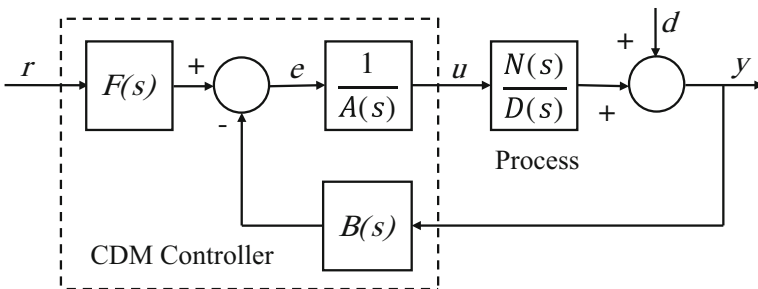


Fig. 10.1 CDM control structure

$$P(s) = \sum_{i=0}^p l_i s^i D(s) + \sum_{i=0}^q k_i s^i N(s) \quad (10.4)$$

The design task is to determine the values of l_i and k_i under the condition that the CDM control system delivers a good (minimum or no overshoot, stable and robust) transient response. In CDM, the above is achieved by means of two design parameters, namely, equivalent time constant (τ) and stability indices (γ_i). These parameters are chosen as follows:

$$\tau = st / (2.5 \sim 3) \quad (10.5)$$

where st is the user defined settling time and

$$\gamma_i = [2.5, 2, 2, \dots] \quad (10.6)$$

The above γ_i values belong to the Manabe's standard form [8]. Several favourable properties of this form can be found in [11]. The designer can change the γ_i values in order to enhance the closed-loop time response.

Using τ and γ_i , a target characteristic polynomial ($P_{\text{target}}(s)$) is framed as

$$P_{\text{target}}(s) = a_0 \left[\left\{ \sum_{i=2}^n \left(\prod_{j=1}^{i-1} \frac{1}{\gamma_{i-j}} \right) (\tau s)^i \right\} + \tau s + 1 \right] \quad (10.7)$$

Equating (10.4) and (10.7), a Diophantine equation [12] is obtained as

$$\sum_{i=0}^p l_i s^i D(s) + \sum_{i=0}^q k_i s^i N(s) = P_{\text{target}}(s) \quad (10.8)$$

By solving (10.8), the coefficients (l_i and k_i) of the CDM controller polynomials are computed. The polynomial $F(s)$ which is defined as the set point filter is chosen to be

$$F(s) = P(s)/N(s)|_{s=0} = P(0)/N(0) = 1/K = k_0 \quad (10.9)$$

10.3 Proposed CDM-PID Tuning Rules

A two-degree-of-freedom PID control system diagram is shown in Fig. 10.2. Here $C_M(s)$ is the main controller chosen in the type of the conventional PID element, $C_F(s)$ is the feed forward controller and $G_P(s)$ is the process to be controlled.

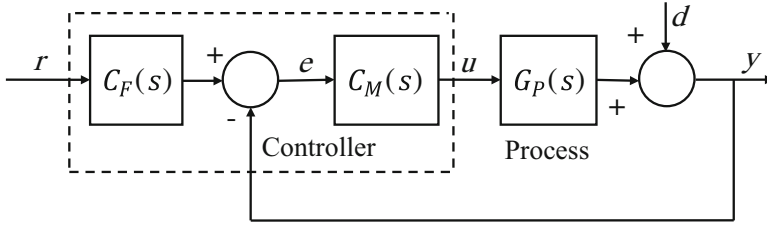


Fig. 10.2 Two-degree-of-freedom PID control system

In this paper, the process is represented or assumed to be approximated as UFOPTD model with the transfer function $G_P(s) = \frac{K}{Ts-1}e^{-Ls}$, where K is the process gain, $T > 0$ is the process time constant and $L > 0$ is the process time delay.

To start with the design, the e^{-Ls} term in the transfer function is approximated using the second-order Taylor denominator approximation technique. After approximation, the $G_P(s)$ is obtained as

$$G_P(s) \approx \frac{K}{(TL^2/2)s^3 + (TL - L^2/2)s^2 + (T - L)s - 1} \quad (10.10)$$

In this work, the controller $C_M(s)$ is chosen as

$$C_M(s) = K_c \left(1 + \frac{1}{T_i s} + T_d s \right) \quad (10.11)$$

where K_c is the proportional gain, T_i is the integral time and T_d is the derivative time.

Using (10.10) and (10.11), the closed-loop characteristic polynomial $P(s)$ is obtained as

$$P(s) = \left(\frac{T_i T L^2}{2} \right) s^4 + \left(T_i T L - \frac{T_i L^2}{2} \right) s^3 + (T_i T - T_i L + K_c K T_i T_d) s^2 + (K_c K T_i - T_i) s + K_c K. \quad (10.12)$$

To derive the PID controller tuning rules based on CDM, it is necessary to frame the CDM target characteristic polynomial $P_{\text{target}}(s)$. Using (10.7), a target characteristic polynomial $P_{\text{target}}(s)$ having the same degree of $P(s)$ is framed as

$$P_{\text{target}}(s) = \left[\frac{a_0 \tau^4}{\gamma_3 \gamma_2^2 \gamma_1^3} s^4 + \frac{a_0 \tau^3}{\gamma_2 \gamma_1^2} s^3 + \frac{a_0 \tau^2}{\gamma_1} s^2 + a_0 \tau s + a_0 \right] \quad (10.13)$$

In Eq. (10.13), the stability index values $\gamma_1 = 2.5$, $\gamma_2 = 2$, $\gamma_3 = 1.25$ are used. The choice of $\gamma_3 = 1.25$ instead of 2 is for enhancing the closed-loop time response. Due

to mathematical constraint, the equivalent time constant τ is considered as a variable to be solved.

Upon equating (10.12) and (10.13), a Diophantine equation is obtained. Symbolically solving the Diophantine equation, the CDM-PID tuning rules and the expression of equivalent time (τ) relating to the UFOPTD process transfer function model are obtained and summarized as.

$$K_c = \frac{\alpha}{25KLT^2}, T_i = \frac{25LT\alpha}{(4T - 2L)^4}, T_d = \frac{10L^3T - 15L^2T^2 + 15LT^3}{\alpha} \quad (10.14)$$

$$C_F(s) = \frac{1}{T_i T_d s^2 + T_i s + 1}, \tau = \frac{25LT}{8T - 4L} \text{ and } \alpha = 32T^3 - 4L^3 + 24L^2T - 23LT^2$$

It is emphasized that these CDM-PID tuning rules for FOPTD process models are new to the control literature. The tuning rules relate the controller parameters directly to the process model parameters. Therefore, it is easy and straightforward to compute the controller parameters once the model parameters are known. The equivalent time constant (τ) and (10.5) help to predict the approximate settling time of the closed-loop response, which is as an added advantage of the proposed tuning rules.

10.4 Simulation Results

The simulation results relating to set point (unit step input) tracking, disturbance (step disturbance of magnitude $d = 0.2$ applied at suitable time at the steady state) rejection and robustness against model uncertainties are provided to demonstrate the potential of the CDM-PID controller utilizing the proposed tuning rules. The results obtained with the proposed tuning rules are compared with MDP-PID [7], IMC-PID [6] and PCS-PID [5] tuning methods. The performance of the controllers put forth in this work is numerically assessed by means of settling time (st_S , st_D) and percentage overshoot ($\%OV_S$, $\%OV_D$). The letters S and D in the suffix denote the set point tracking and disturbance rejection, respectively. Besides, the sum of the total variation of the control signal which evaluates the control signal (u) usage is calculated as $TV_{\text{sum}} = TV_{\text{set point tracking}} + TV_{\text{disturbance rejection}}$. For a well-tuned controller, TV_{sum} should be as small as possible [13].

10.4.1 Set Point Tracking and Disturbance Rejection

The UFOPTD transfer function $G_{P1}(s) = \frac{1}{s-1} e^{-0.4s}$ given in [6, 7] is considered. Substituting the model parameters of $G_{P1}(s)$ in the tuning rules presented in (10.14), the CDM-PID controller parameters $K_c = 2.6384$, $T_i = 2.5162$, $T_d = 0.1607$, the equivalent time constant $\tau = 1.5625$ and $C_F(s) = 1/(0.4044s^2 + 2.5162s + 1)$ are

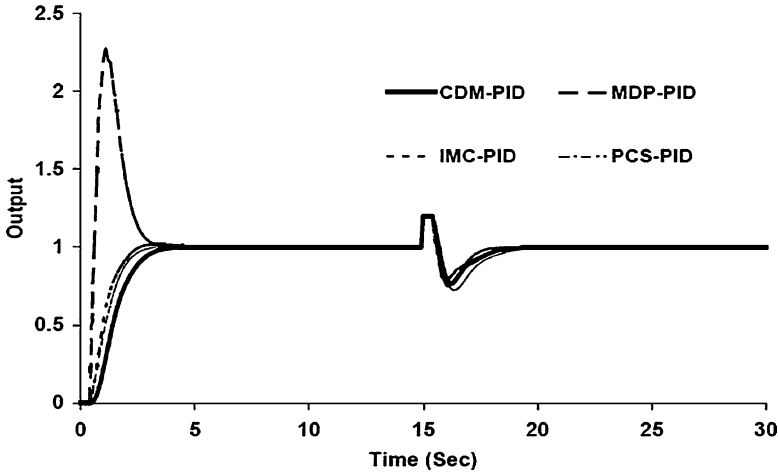


Fig. 10.3 Closed-loop responses for $G_{P1}(s)$

Table 10.1 Performance measures for $G_{P1}(s)$

Measures	CDM-PID	MDP-PID	IMC-PID	PCS-PID
st_S	4.7	5.0	6.0	4.5
$\%OV_S$	Nil	126.9	1.73	Nil
st_D	5.4	3.8	5.0	7.0
$\%OV_D$	114.1	126.2	109.3	137.3
TV_{sum}	22.4	59.9	43.6	4.6

determined. By substituting this τ in (10.5), it is predicted that the closed-loop response has a settling time of around 4.68 s. For comparison, the MDP-PID ($K_c = 0.2222$, $T_i = 0.2011$, $T_d = 0.0561$, $\alpha = 2.1935$ and $\beta = 0.05$) parameters, IMC-PID ($K_c = 2.5$, $T_i = 2.67$, $T_d = 0.17$ with set point weighting coefficient $\varepsilon = 0.2$) parameters and PCS-PID controller parameters ($K_c = 4.75$, $K_{P1} = 0.625$, $T_{i1} = 0.2$, $K_{P2} = 2.0799$, $T_{i2} = 3.3898$, $\alpha = 0.2$ and $\beta = 0.0863$) are used. The closed-loop simulations are performed. From Fig. 10.3, it is apparent that the CDM-PID and PCS-PID controller gives a smooth set point tracking response with zero overshoot and shorter settling time. The disturbance rejection capability of the CDM-PID is better than PCS-PID and close to those achieved by other two controllers. The performance measures tabulated in Table 10.1 shows that the settling time of CDM-PID is very close to the value predicted. The TV_{sum} of CDM-PID is also less compared to those of MDP-PID and IMC-PID. It indicates that the CDM-PID with the proposed tuning rules supplies a good quality control with a small magnitude of control signal.

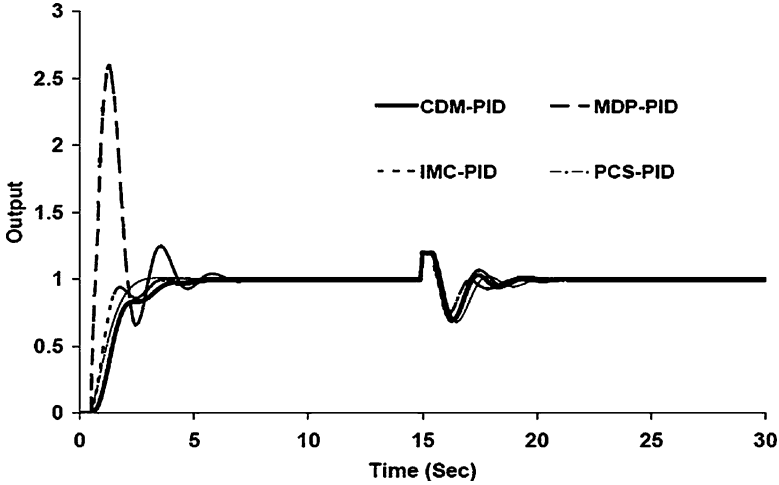


Fig. 10.4 Closed-loop responses for $G_{P1}^*(s)$

Table 10.2 Performance measures for $G_{P1}^*(s)$

Measures	CDM-PID	MDP-PID	IMC-PID	PCS-PID
st_S	7.7	9.2	7.9	5.5
$\%OV_S$	Nil	160.4	Nil	0.8
st_D	7.8	7.7	7.6	8.5
$\%OV_D$	149.0	159.0	138.6	159.1
TV_{sum}	22.9	62.2	56.5	5.2

10.4.2 Robustness Test

The robustness test is performed by perturbing the model parameters of the UFOPTD transfer function $G_{P1}(s)$. Here, the model parameters of $G_{P1}(s)$ are simultaneously [6] perturbed to +20% of its original value. Using the perturbed model $G_{P1}^*(s) = \frac{1.2}{1.2s-1} e^{-0.48s}$ but the PID controller settings based on the $G_{P1}(s)$, a closed-loop simulation is performed.

The results (Fig. 10.4 and Table 10.2) indicate that all the four PID controllers achieve a robust response. However, the CDM-PID controller yields better set point tracking response with short settling time and nil overshoot. Compared to MDP-PID and PCS-PID, the disturbance rejection response of CDM-PID has shorter settling time with a minimum overshoot. The TV_{sum} of CDM-PID is still less than those of MDP-PID and IMC-PID controllers. It confirms that the CDM-PID with the proposed tuning rules supplies a robust response with a small magnitude of consistent control signal.

Table 10.3 L/T ratio stability region

Measure	CDM-PID	MDP-PID	IMC-PID	PCS-PID
L/T	≤ 1.7	≤ 1.2	≤ 0.5	≤ 1.0

10.4.3 Stability Analysis

In this section, the analysis is carried out by investigating the effect of the ratio of the time delay (L) to the time constant (T), i.e. L/T ratio on the closed-loop stability [14]. At different values of L/T , the CDM-PID controller and other controllers are calculated. Based on the closed-loop simulation results, the L/T ratio stability region of CDM-PID is compared with those of MDP-PID, IMC-PID and PCS-PID in Table 10.3. This table clearly indicates that, compared to other PID controllers, the CDM-PID with the proposed rule is superior and offers stability for the largest L/T ratio.

10.5 Conclusion

This paper has proposed CDM-based two-degree-of-freedom PID controller tuning rules for the unstable first-order plus time delay processes. The explicit tuning rules depending on the UFOPTD model parameters have been derived by using the CDM concept and the second-order Taylor denominator approximation. The derived rules are simple and straightforward to use. The closed-loop performance of the systems with CDM-PID controller employing the proposed tuning rules is investigated through the simulations. Investigations reveal that the CDM-PID controller shows better performances compared to other PID tuning methods. As evidenced by its performances, it is concluded that the CDM-PID controller with the proposed tuning rules gives promising results for unstable first-order plus time delay processes.

References

1. Srivastava S, Pandit VS (2016) A PI/PID controller for time delay systems with desired closed loop time response and guaranteed gain and phase margins. *J Process Control* 37:70–77
2. Vanavil B, Krishna Chaitanya K, Seshagiri Rao A (2015) Improved PID controller design for unstable time delay processes based on direct synthesis method and maximum sensitivity. *Int J Syst Sci* 46(8):1349–1366
3. Arvanitis KG, Soldatos AG, Boglou AK, Bekiaris-liberis NK (2009) New simple controller tuning rules for integrating and stable or unstable first order plus dead-time processes. In: *Proceedings of 13th WSEAS international conference on systems*, pp 328–337
4. Panda RC (2009) Synthesis of PID controller for unstable and integrating processes. *Chem Eng Sci* 6(12):2807–2816
5. Ajmeri M, Ali A (2015) Two degree of freedom control scheme for unstable processes with small time delay. *ISA Trans* 56:308–326

6. Wang Q, Lu C, Pan W (2016) IMC PID controller tuning for stable and unstable processes with time delay. *Chem Eng Res Des* 105:120–129
7. Ravi Kishore C, Padma Sree R (2018) Design of PID controllers for unstable systems using multiple dominant pole placement method. *Indian Chem Eng* 60(4):356–370
8. Manabe S (1998) Coefficient diagram method. In: *Proceedings of 14th IFAC symposium on automatic control in aerospace*, Seoul, pp 199–210
9. Hamamci SE, Bhaba PK, Somasundaram S, Karunanithi T (2007) Two degree of freedom transient response control for unstable processes. In: *Proceedings of 5th international conference on trends in industrial measurements and automation*, India, pp 145–150
10. Hamamci SE (2014) A new PID tuning method based on transient response control. *Balkan J Electr Comp Eng* 2(3):132–138
11. Manabe S, Kim YC (2000) Recent development of coefficient diagram method. In: *Proceedings of the 3rd Asian control conference*, Shanghai
12. Kucera V (1993) Diophantine equations in control—a survey. *Automatica* 29(6):1361–1375
13. Skogestad S (2003) Simple analytic rules for model reduction and PID controller tuning. *J Process Control* 13:291–309
14. Jung CS, Song HK, Hyun JC (1999) A direct synthesis tuning method of unstable first order plus time delay processes. *J Process Control* 9:265–269

Chapter 11

Real-Time Energy Management System for Solar-Wind-Battery fed Base Transceiver Station



W. Margaret Amutha and V. Rajini

Abstract This chapter proposes an intelligent energy management system which integrates solar and wind energy systems with battery backup for making best use of their operating characteristics and obtain better efficiency. Energy management system is programmed for maintaining the energy sustainability in solar–wind renewable energy systems, constant power at point of common coupling, regulating the reference currents based on instantaneous power delivered by the sources and load demand. A controller-based state of charge estimation, charging and discharging of batteries to suitably source or sink the input power according to the load demand is also presented. The simulation of the proposed energy management controller is done using Matlab/Simulink. The simulation results are encouraging in reliability and stability perspective. Experimental results based on solar system with 1.5 kW peak power, wind system with 1.4 kW peak power and 48 V, 200 Ah lead–acid batteries with embedded controller validate the theoretical approach. It is observed that the proposed controller provides constant power at the point of common coupling, thus meets the load demand without interruption.

Keywords Base transceiver station · Energy management system · Embedded controller · Point of common coupling · State of charge estimation · Solar · wind · battery system

W. M. Amutha (✉)

Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: margaret@karunya.edu

V. Rajini

SSN College of Engineering, Chennai, India
e-mail: rajiniv@ssn.edu.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_11

Abbreviations

BTS	base transceiver station
DOD	depth of discharge
PCC	point of common coupling
SOC	state of charge

11.1 Introduction

Though there are concerns due to the intermittent nature, mismatch and consumption in power, renewable sources seem to be a key solution in reducing greenhouse gases, CO₂ emissions and fuel consumption [1]. DC–DC converters are essential as most renewable energy-powered systems, input power delivered, output load demand or instantaneous change in both are not exactly identical with each other at any time instant. So, this work uses a single-fused cuk–buck DC–DC converter [2] as it uses reduced number of components that are compact and efficient [3–7]. The development of renewable energy systems is hindered partly due to the lack of feasible energy storage solutions. In most cases, the seasonal variation and intermittent nature of renewable energy sources leads to the deployment of energy storage solutions [8–11]. So, the energy storage is an important element in a hybrid renewable energy system, simply to fill the gaps of supply or during unexpected or irregular load consumption [12–15].

The energy storage system stores surplus energy during the periods when the energy generation is more than the load demand and releases this energy when the load demand is more than the energy generation [16–19]. The energy storage also acts as an energy buffer by compensating the imbalances found between the energy generation and the demand and it improves the system's performance so that any power interruption within the telecom networks can be mitigated. It is also necessary to prioritize the sources for consumption, manage the power flow from sources to load, state of charge estimation, charging/discharging of batteries, instantaneous load demand and energy management. The aim of this chapter is to present a controller which is able to self-regulate the set of renewable energy sources, backup system according to instantaneous load demand. The controller will choose the most economical source of energy from the available sources and that source will drive the attached load at the output. The controller also decides when to use the sources or when to use the energy stored in the battery.

This chapter is organized as follows. The first section is a short review of the literature. In the second block diagram, a system description and simulation result of single stage cuk–buck converter is presented. Third section deals about the

energy management controller, online state of charge (SOC) estimation, charging and discharging of batteries. Fourth section contains the real-time implementation of solar-wind system-based cuk-buck converter with energy management controller and the results. Finally, the findings of the investigations are highlighted in the conclusion.

11.2 Block Diagram and System Description

Figure 11.1 shows the block diagram of a single cuk-buck fused converter with energy management system.

From the architecture, the controller receives the instantaneous voltage and current of input sources V_S , V_W , I_S , I_W , battery voltage V_B , battery current I_B , state of charge (SOC). According to the received signals, controller fixes the width of the duty cycles d_S , d_W , d_{bc} and d_{bd} to feed in the required power to the load. The topology of cuk-buck converter with energy management controller is shown in Fig. 11.2. Here cuk converter is used to process solar energy whereas buck converter is used to process wind energy. The circuit is simulated in Matlab/Simulink with the parameters, L_1 of 158 μH , L_2 of 84.16 μH , C_1 of 2906 μF , C_2 of 1132 μF , resistive load of 1.92 Ω and switching frequency of 20 kHz for its investigation. It is desired to obtain -48 and 23 A from inputs of 65 and 55 V

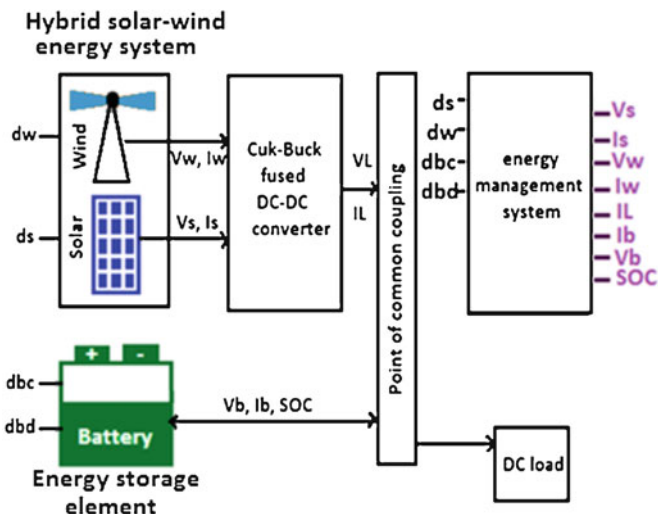


Fig. 11.1 Block diagram of proposed system

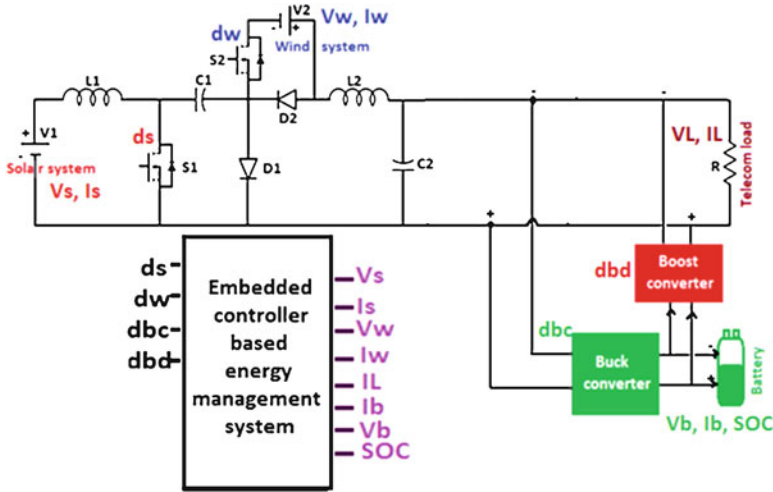


Fig. 11.2 Topology of proposed system

11.2.1 Flow Diagram of the Control Logic

The flow diagram describing the control logic is shown in Fig. 11.3. The controller controls the cuk–buck fused converter by updating the operating current of the solar and wind sources via varying the duty cycle of fused converter in a variable step manner till it satisfies the load demand. The controller generates the duty cycle to develop point of common coupling voltage $V_{PCC} = -48 \text{ V}$ as given in Eqs. (11.1) and (11.2) for solar and wind sources:

$$d_{Sref} = \frac{V_{PCC} - V_S}{V_{PCC}} \tag{11.1}$$

$$d_{Wref} = \frac{V_{PCC} - V_W}{V_{PCC}} \tag{11.2}$$

where V_{PCC} is the voltage at the point of common coupling, V_S is the input voltage of cuk converter connected to solar system and V_W is the input voltage of buck converter connected to wind system.

When the point of common coupling voltage is set at -48 V , the controller compares the load current with the solar current and wind current:

$$I_{sref} + I_{wref} > I_L \tag{11.3}$$

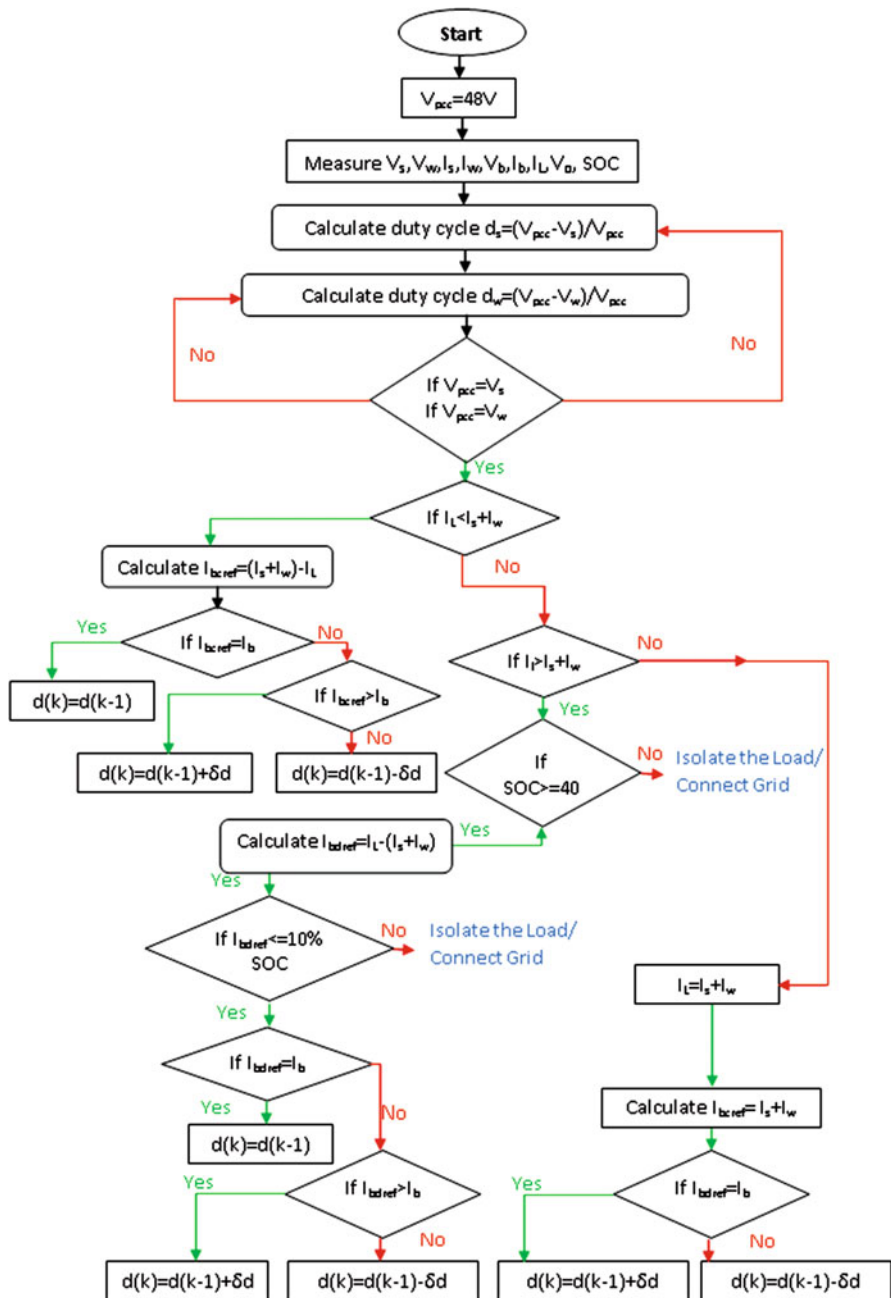


Fig. 11.3 Flow diagram of the control logic

If load current follows Eq. (11.3), the reference charging current of the battery proportional to excess power available at point of common coupling is estimated as follows:

$$(I_{\text{sref}} + I_{\text{wref}}) - I_L = I_{\text{BCref}} \quad (11.4)$$

When actual I_{BC} is equal to the reference I_{BC} of the battery, duty cycle (d_{bc}) of buck converter charging the battery is maintained the same by keeping the same reference voltage. When load demand is higher, battery is discharged along with solar and wind to supply the load demand with the duty cycle of d_{bd} . Controller is programmed to discharge the battery only if the state of charge of the battery is more than 40%. The discharging reference current is calculated by

$$I_L - (I_{\text{sref}} + I_{\text{wref}}) = I_{\text{BDref}} \quad (11.5)$$

If the load demand is still higher than the sum of power generated by solar and wind system and peak dischargeable power of the battery, the load should be disconnected from the power supply.

11.3 Backup Facilities

11.3.1 Battery Specifications

Battery is used as the external smoothening agent to sink or source the power based on the instantaneous load demand. The lead–acid batteries are preferred for telecom applications as the maintenance and the initial costs are less. The rate of charging and the discharging of the battery are done based on the standard specifications of the battery handbook. The lead–acid battery handbook illustrates that the charging current of the battery should be less than $0.1C_B$.

$$\text{Battery charging} = 0.1 * C_B \quad (11.6)$$

For a 200 Ah battery the charging current should not exceed:

$$I_{\text{Battery charging}} = 0.1 * 200 = 20 \text{ A} \quad (11.7)$$

The capacity of the battery (C_B) needed for delivering the power of 1.2 kW (telecom load) at the battery voltage of 48 V and 85% of battery efficiency shall be calculated as

$$C_B = \frac{1200 * 6 * 1}{0.8 * 48 * 0.85} = 220.58 \text{ Ah} = 200 \text{ Ah} \quad (11.8)$$

The maximum battery discharge current at the output of the converter to deliver a power of 1.2 kW at the V_{Battery} of 48 V and the battery efficiency 0.85% is

$$I_{\text{Battery discharge}} = \frac{1200}{48 * 0.85} = 29.4 \text{ A} = 30 \text{ A} \tag{11.9}$$

For the simulation purpose, 2 numbers of 24 V batteries are considered. 12 V × 4 numbers of batteries = 48 V. So, 4 numbers of parallel connected 200 Ah batteries are considered for hardware implementation.

11.3.2 SOC Estimation of Battery

The battery is set at 40% state of charge (SOC) as depth of discharge (DOD) to about 70–80% of its capacity shall damage the battery.

Tables 11.1 and 11.2 show the terminal voltage at various charging currents and discharging currents for 24 V lead–acid battery. The charging and discharging process takes place depending on the current flow direction. The battery current

Table 11.1 Terminal voltage at various charging current

	Voltage at (C/50)	Voltage at (C/30)	Voltage at (C/20)	Voltage at (C/10)
10	21	21.4	22	22.5
20	22	22.1	22.21	22.52
30	22.2	22.3	22.4	22.63
40	22.42	22.45	22.5	22.92
50	23.1	23.2	23.25	23.48
60	23.3	23.4	23.5	23.52
70	23.5	23.61	23.63	23.74
80	23.6	23.7	23.9	24
90	24.24	24.12	24.13	24.24
100	24.1	24.15	24.2	24.46

Table 11.2 SOC Vs terminal voltage at various discharging current

	Voltage at (C/75)	Voltage at (C/50)	Voltage at (C/30)	Voltage at (C/20)	Voltage at (C/10)
10	21.2	21.22	21.23	22.6	23
20	22	22.7	22.8	22.82	23.2
30	22.4	22.9	23	23.2	23.3
40	22.6	23.2	23.3	23.4	23.6
50	22.8	23.3	23.4	23.5	23.7
60	22.9	23.5	23.6	23.9	24.2
70	23	23.9	24.2	24.4	24.4
80	23.2	24.2	24.3	24.45	24.8
90	23.3	24.3	24.4	24.6	25
100	23.6	23.4	24.5	24.8	25.2

direction and voltage are the inputs of the online SOC estimation. The SOC estimation for the terminal voltage 24 V and the current of 15 A during charging process can be estimated by the following Eq. (11.10):

$$\text{SOC} = \left(\frac{(B - A) * (V - V_1)}{(V_2 - V_1)} + A \right) \quad (11.10)$$

where V is the battery voltage, V_1 is the voltage < the terminal voltage in the corresponding column of given current, V_2 is the voltage > the terminal voltage in the corresponding column of given current, A is the V_1 row SOC reference and B is the V_2 row SOC reference.

If the battery current or voltage lies between any two values specified in Table 11.1 either row or column wise, the SOC of battery for that voltage and current is determined using Eq. (11.11), using the row-wise neighbouring values of voltage between which the battery current falls. For the estimation of charging current 15 A, the battery current lies in between the columns ($C/10$) and ($C/20$), then a new column should be created virtually and then SOC needs to be estimated using the following formula:

$$V_1 = \left(\frac{(22.5 - 22) * (I - \frac{C}{20})}{(\frac{C}{10} - \frac{C}{20})} + 22 \right) \quad (11.11)$$

Likewise, V_2, V_3 till V_{10} are calculated. Then calculate SOC using Eq. (11.10). Similarly, SOC estimation for discharging process is calculated using the above procedure and discharge data.

11.3.3 Performance of SOC Estimator

State of charge of the battery at various currents during charging mode is estimated using the above-mentioned algorithm. Figure 11.4 shows the actual state of charge of the battery during charging mode. As given in Table 11.1, it verifies SOC estimation during charging mode of the battery. State of charge of the battery at various currents during discharging mode is also estimated using the same algorithm. Figure 11.5 shows the actual state of charge of the battery while discharging. As given in Table 11.2, it verifies SOC estimation during discharging mode of the battery.

11.3.4 Performance of Controller

Simulation is carried out for various seasonal changes like non-sunny–non-windy days, sunny–windy days, sunny days and windy days by disconnecting or

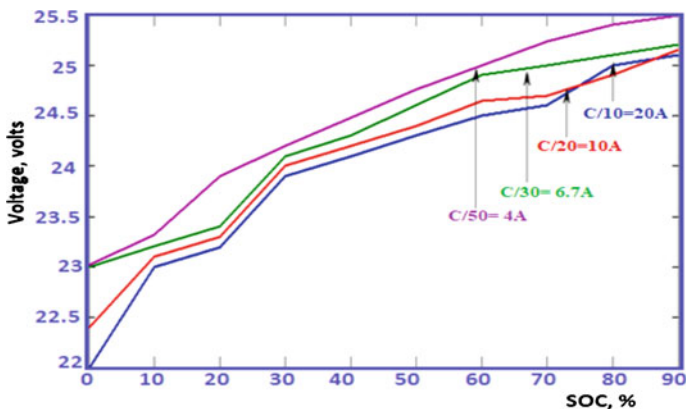


Fig. 11.4 Battery SOC (charging mode)

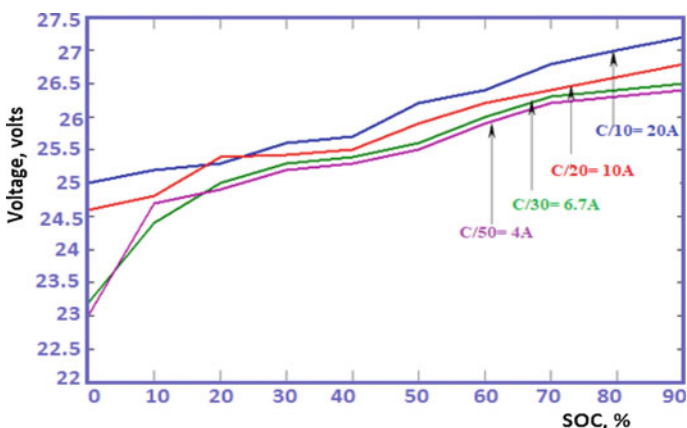


Fig. 11.5 Battery SOC (discharging mode)

connecting the input solar and wind sources. As shown in Fig. 11.6, the change in the load voltage and load current during different seasons is compensated by energy stored in the battery. The simulation results verify the same. The voltage and current output of the converter with and without controller is shown in Fig. 11.7.

11.4 Hardware Description

Energy management system for solar-wind hybrid system with the cuk-buck converter is implemented in real time. The topology is tested under the same seasonal conditions and input power. Research Centre rooftop, SSN College of Engineering, Kalavakkam in Chennai district in the Indian state of Tamil Nadu is

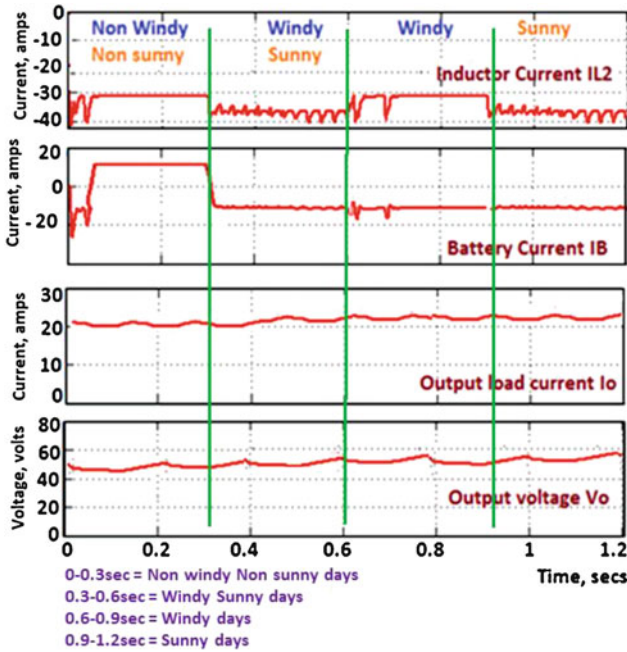


Fig. 11.6 Performance of the controller

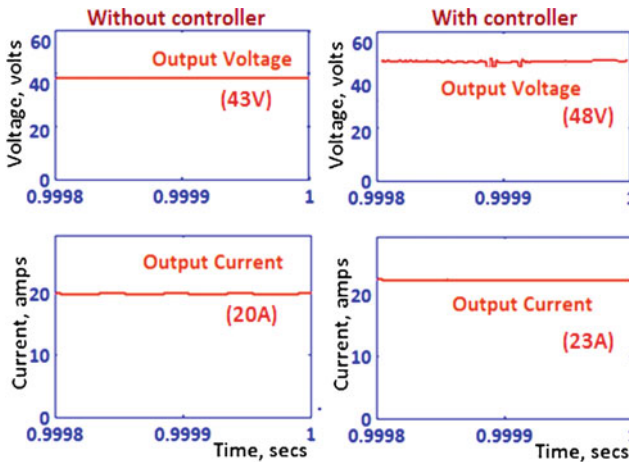


Fig. 11.7 Voltage and current waveforms

considered. Figure 11.8 shows the view of single-stage cuk–buck–fused converter setup with embedded controller. The six panels are connected in three parallel paths with two panels in series with each path. Based on the criteria, wind turbine of

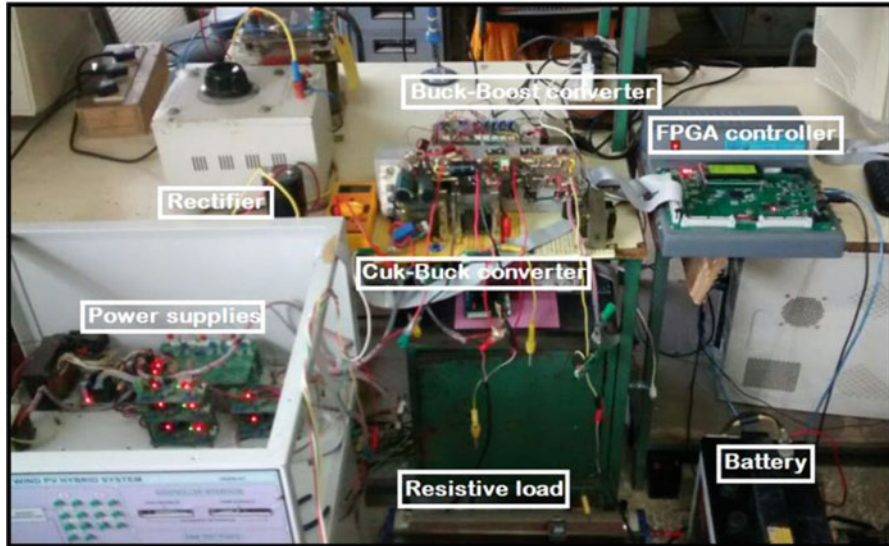


Fig. 11.8 Cuk-buck-fused converter

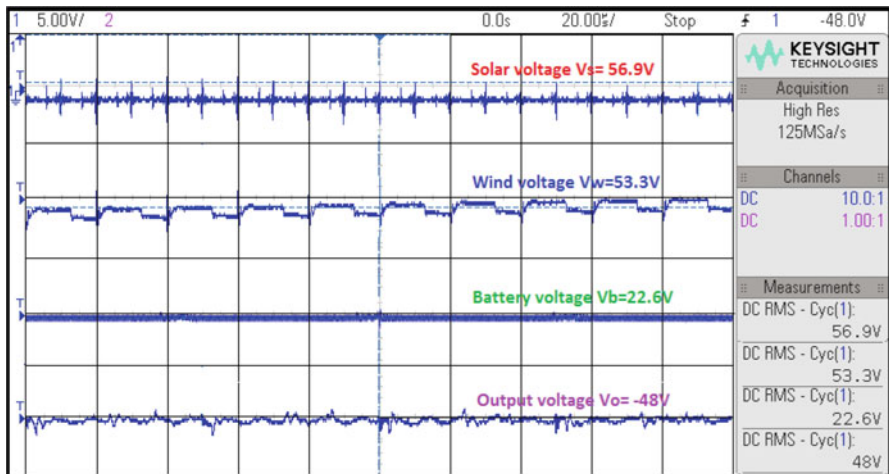


Fig. 11.9 V_i and V_o

1.4 kW, the rated wind speed of 15 m/s, the cut in speed of 3 m/s are considered. It has permanent magnet synchronous generator type with 48 V dc rectifier circuit. Four numbers of 12 V batteries are connected in series to get 48 V. Load is connected with the point of common coupling through the battery bank.

Figures 11.9 and 11.10 show the solar, wind and battery input voltage and currents, output load voltage and current of cuk-buck converter. From the figures,

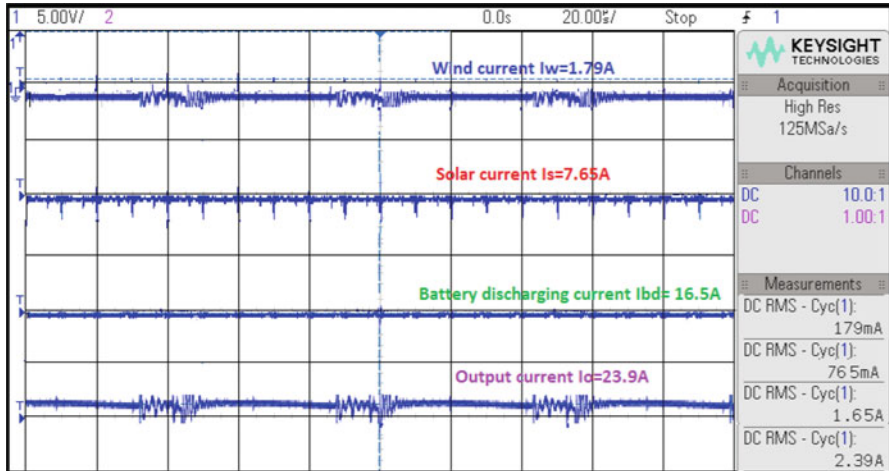


Fig. 11.10 I_i and I_L

it is clear that the produced supply currents from hybrid renewable energy sources ($7.65\text{ A} + 1.79\text{ A} = 9.44\text{ A}$) are not able to meet the required load current. So, battery current ($I_b = 16.5\text{ A}$) supports the required load current.

11.5 Conclusions

Thus solar–wind-based cuk–buck-fused converter with embedded controller is implemented which provides uninterrupted supply. An embedded controller based on state of charge estimation, charging and discharging of batteries to suitably source or sink the input power according to the load demand is also presented. Experimental results based on solar system with 1.5 kW peak power, wind system with 1.4 kW and 48 V, 200 Ah lead–acid batteries validate the theoretical approach.

References

1. Saravanan K, Habeebullah Sait H (2013) Multi input converter for distributed renewable energy sources. *Int J Emer Trends Electr Electron* 2(1):51–58
2. Mangu B, Kiran Kumar K, Fernandes BG (2012) Efficiency improvement of solar-wind based dual input converter for telecom power supply. *IEEE conference*, pp 1–6
3. Li Y, Ruan X, Yang D, Liu F, Tse CK (2010) Synthesis of multiple-input dc-dc converters. *IEEE Trans Power Electron* 25(9):2372–2385
4. Vivekanandan S, Kamalakannan P (2016) Modelling and simulation of multi input boost converter for renewable energy applications. *Int J Sci Eng Appl Sci* 2(2)

5. Yalamanchili KP, Ferdowsi M (2005) Review of multiple input DC-DC converters for electric and hybrid vehicles. In: Proceedings of the IEEE vehicle power and propulsion conference 7–9, pp 160–163
6. Choung SH, Kwasinski A (2005) Multi-input dc-dc converter topologies comparison. In: Proceedings of the IEEE industrial electronics conference (IECON), pp 2359–2364
7. Olallaa C, Queinneca I, Leyvac R, El Aroudic A (2011) Robust optimal control of bilinear dc-dc converters. *Control Eng Pract* 19(7):688–699
8. Czarkowski D, Kazimierczuk MK (1993) Energy conservation approach to modelling pwm dc-dc converters. *IEEE Trans Aerosp Electron Syst* 29(3):1059–1063
9. Eichhorn T (2008) Boost converter efficiency through accurate calculations. *Power Electr Technol* 34:30–35
10. Carusone TC, Johns DA, Martin KW (2012) Chapter 1: Integrated circuit devices and modelling. In: *Analog integrated circuit design*, 2nd edn. Wiley, Hoboken, pp 10–13; 62–64
11. Li CH, zhu XJ, cao G y, Sui S, Ruohu M (2009) Dynamic modelling and sizing optimization of stand-alone photovoltaic power systems using hybrid energy storage technology. *Renew Energy* 39:815–826
12. Markvart T (1996) Sizing of hybrid photovoltaic-wind energy systems. *Sol Energy* 57(4):277–281
13. Shezana SKA, Dasb N, Mahmudula H (2016) Techno-economic analysis of a smart-grid hybrid renewable energy system for Brisbane of Australia. In: *Energy Procedia*, 1st international conference on energy and power, ICEP2016, vol 110, pp 340–345
14. Shezan SA, Julai S, Kibria MA, Ullah KR, Saidur R, Chong WT, Akikur RK (2016) Performance analysis of an off-grid wind-PV (photovoltaic)-diesel-battery hybrid energy system feasible for remote areas. *J Clean Prod* 125:121–132
15. Chinchilla M, Arnaltes S, Burgos JC (2006) Control of permanent-magnet generators applied to variable-speed wind-energy systems connected to the grid. *IEEE Trans Energy Convers* 21(1):130–135
16. Siemens (2011) Energy management and energy optimization in the process industry, September 2011, Siemens AG, pp 1–28
17. Bianchi MVA (2011) Challenges and Opportunities to achieve 50% energy savings in homes. *Nat Lab*, pp 1–104
18. Cirrincione M, Cossentin M, Gaglio S, Hilaire V, Koukam A, Pucci M, Sabatucci L Vitale G (2009) Intelligent energy management system, *IEEE explore*, July 2009
19. Thangavel S, Saravanan S (2014) A simple power management scheme with enhanced stability for a solar PV/wind/fuel cell fed standalone hybrid power supply using embedded and neural network controller. *J Electr Eng Technol* 9(5):1454–1470

Chapter 12

IOT-Based Adaptive Protection of Microgrid



O. V. Gnana Swathika and K. T. M. U. Hemapala

Abstract Microgrid is a localized power network of connected loads and energy distribution sources that may operate autonomously with respect to the main central grid. Conventional protection schemes are not suitable for grid-connected and islanded modes of operation of microgrid. This chapter proposes an adaptive protection centre (APC) that is capable of real-time monitoring of the microgrid, fault identification and graph algorithms-aided shortest path identification for fault clearance. The Internet of Things (IOT)-based APC is capable of displaying system status and load characteristics in a remote system. The proposed APC is validated and tested for functionality on the hardware prototype of eight-bus microgrid network.

Keywords Adaptive Protection · Microgrid · IOT

Abbreviations

APC	Adaptive protection CENTRE
CT	Current transformer
DG	Distributed generators
IOT	Internet of Things
MPC	Microgrid protection CENTRE
PCC	Point of common coupling
UG	Utility grid

O. V. G. Swathika (✉)
VIT University, Chennai, Tamil Nadu, India
e-mail: gnanaswathika.ov@vit.ac.in

K. T. M. U. Hemapala
University of Moratuwa, Moratuwa, Sri Lanka

12.1 Introduction

Microgrid is a localized power network of connected loads and energy distribution sources that may operate autonomously with respect to the main central grid [1–3]. A central monitoring system records the status of all the connected distributed generators (DGs), loads and utility grid. Protection strategies assisted with communication systems are a common solution to microgrid protection [4]. This chapter presents an adaptive protection scheme for an eight-bus microgrid network by realizing a microgrid protection centre (MPC) that is capable of identifying the fault location and fault clearance. The status of the system is continuously monitored by a remote monitoring system via Internet of Things (IOT), which displays the real-time current values of the buses. It also does fault detection and rectification in the microgrid system. It employs Graph Theory Algorithms at MPC to find the shortest path from the faulted point to its nearest operating source. The system status and load characteristics are displayed on a remote device with the help of IOT.

12.2 Proposed IOT-Based Microgrid Protection System

The methodology followed by the proposed system is given in Fig. 12.1. A hardware prototype of the eight-bus microgrid test system is realized and monitored using the MPC. A fault is initiated at a node of the network and is detected by the MPC. Based on the integrated algorithm, the shortest path for fault clearing is identified and realized. The microgrid status and load parameters are updated to a remote monitoring system via IOT, as well as displayed at the local display unit.

12.2.1 Methodology

As the system turns ON, the user is required to choose from the protection scheme options. Once a valid scheme is selected, the control system identifies the active nodes of the microgrid and the node status is updated into the system memory. The active nodes are mapped into a network and the system initiates load current monitoring. Various input ports read the load parameters (load current and status) and updates the values to a local database (internal memory – RAM). The load current values are compared to the reference limits and fault detection sequence is initiated. If a fault is detected at any node, the controller triggers the corresponding nodes and disconnects them from the system and the load parameters are reread for values and status. In case, there is no fault in the system, the entire algorithm executes in a cyclic manner.

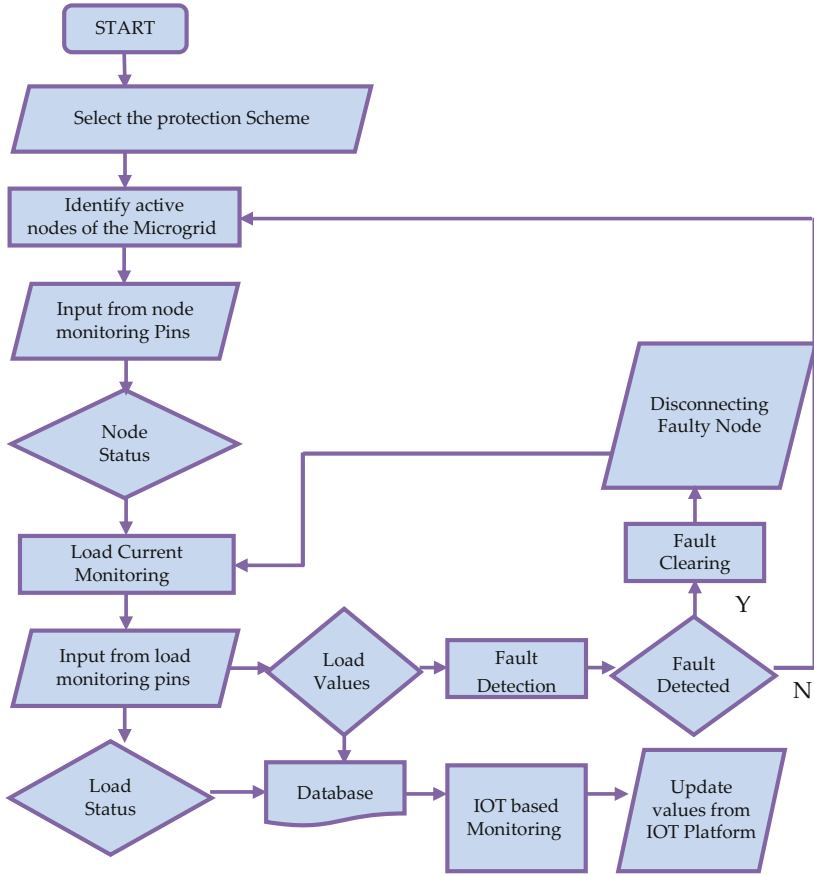


Fig. 12.1 Methodology

12.2.2 Modules

The different modules involved are test system, relay network, Microgrid Protection Centre (MPC) and Ethernet Shield.

Test System The eight-bus microgrid test system shown in Fig. 12.2 is considered for analysis. The utility grid (UG) supplies the load requirement of the microgrid in grid-connected mode. The grid is rated at 34.5 kV. There are two zones in the network that are connected to the UG through the Point of Common Coupling (PCC) named as CB0. The first zone comprises circuit breakers from CB1–CB4 and the second zone consists of CB5–CB7. Each circuit breaker caters to a resistive load of 2 MW. The system follows an adaptive protection Scheme. CO-7 overcurrent relays are used, which trip following a CTI of 0.3 s. The maximum current allowed in the system is 600 A.

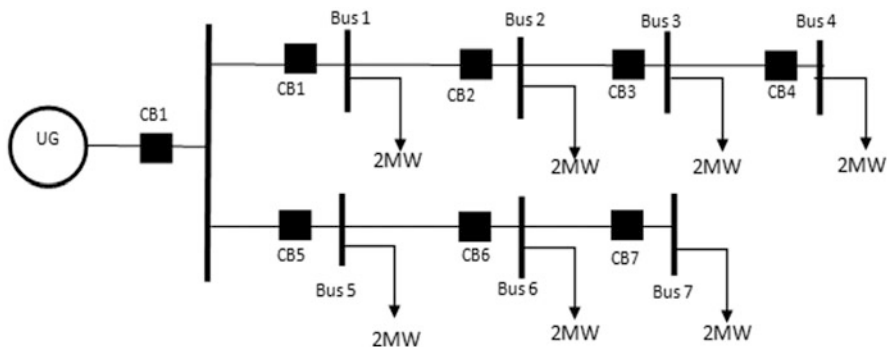


Fig. 12.2 Sample microgrid network

Relay Network Relays compatible with Arduino are chosen to represent the system. This module gets an input from the MPC and exhibits adaptive protection by tripping corresponding relays. The system consists of a network of seven overcurrent relays. The Simulink Model depicting the prototype relay network is realized.

Microgrid Protection Centre The MPC comprises the Arduino ATmega 2560 (containing the database and tripping code) and a display (online IOT dashboard or Laptop). The objective of the MPC is to monitor the microgrid and capture feeder currents at any instant of time. It compares the captured data with the values existing in the database and displays fault details (if any). The status of the system is analysed and stored on a remote monitoring system. The communication is established through IOT.

Arduino Ethernet Shield W5100 The Ethernet Shield W5100 is based on WIZnet W5100, and a single chip Internet enabled 10/100 Ethernet controller. The objective of the IC is to implement Internet connectivity without OS.

12.3 Design Approach

The microgrid test system is simulated in MATLAB Simulink to obtain the nominal currents by performing load flow analysis. By using the values obtained from the simulation, “top” values are calculated and the overcurrent relay coordination is achieved using the Prims–Johnsons Graph Theory Algorithm.

The system status, that is, active nodes, protection scheme, active loads and load currents of the system, is monitored and updated on the controller (Arduino Mega) memory. This data is sent to a remote monitoring system with the help of IOT. The remote monitoring system displays the real-time values of currents on the different buses. If there is a fault in the system, a visual indication will be triggered on the remote system. The display for the remote monitoring is carried on an open source

Table 12.1 Technical specification

Components	Rating
Arduino mega (Atmega2560)	Operating voltage: 5 V
	Operating current: 50 mA
	Input voltage: 7–12 V
	Clock speed: 16 MHz
	Per pin output rating: 40 mA
Ethernet shield W5100	Operating voltage: 5 V
	Operating current: 50 mA
	Input voltage: 5 V
	Clock speed: 25 MHz
LEDs	0.8 V
Relays	Coil: 5 V
	Output: 5 A, 120 V (DC)
Variable loads (resistive)	1 k Ω , 0.5 W
Fixed load (resistive)	220 Ω , 0.5 W
Jumper cables	5 V, 100 mA

platform, Cayenne. There are three phase loads, generators (as DG) and distribution-line impedances. Microgrid test system specifications are as in Fig. 12.2. The parameters of the test system are scaled down for testing. Table 12.1 shows the technical specification of the prototype. The objective of the MPC is to monitor the microgrid parameter and status, capture feeder currents at any instant of time based on the input and relays the data to the display unit. It compares the captured data with the values existing in the database and displays fault details. The status of the system is then monitored using IOT on a remote system. This is enabled by the use of Ethernet Shield W5100 compatible with ATmega2560. The steps involved are computing simulation results and calculations, hardware implementation and obtaining hardware results. The currents computed from the simulation are used to select the current transformer (CT) ratios of the system. In this paper, 5 V relays are used. The pin connections are shown in Fig. 12.3.

The load is connected to the NO contact of the relay. In normal operating conditions, the relay is activated by initializing a control signal in the code (from the MPC) and the load is connected to the grid. The signal is sent by the MPC to Pin 2, which is connected to the relay's control coil. When the MPC detects a fault, another signal is sent to the control pin and the control coil de-energizes the electromagnet that open circuits the load circuit, thus restricting current flow in the load branch.

The load branch consists of a 1 K Ω variable resistive load, an LED and a fixed load of 220 Ω . The analogue pins of the Arduino are connected to the fixed loads. These connections enable the measurement of the voltage equivalent of current flowing through the fixed loads. These values are used by the MPC for fault detection. The digital pins of the Arduino are connected to Pin 1 of the relays. Two digital pins are connected to each relay, one for sensing the status of the relay and another for controlling relay.

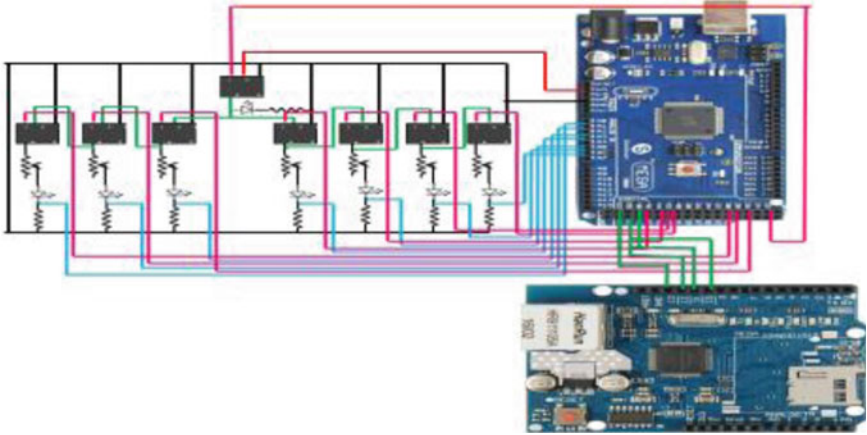


Fig. 12.3 H/W prototype connections

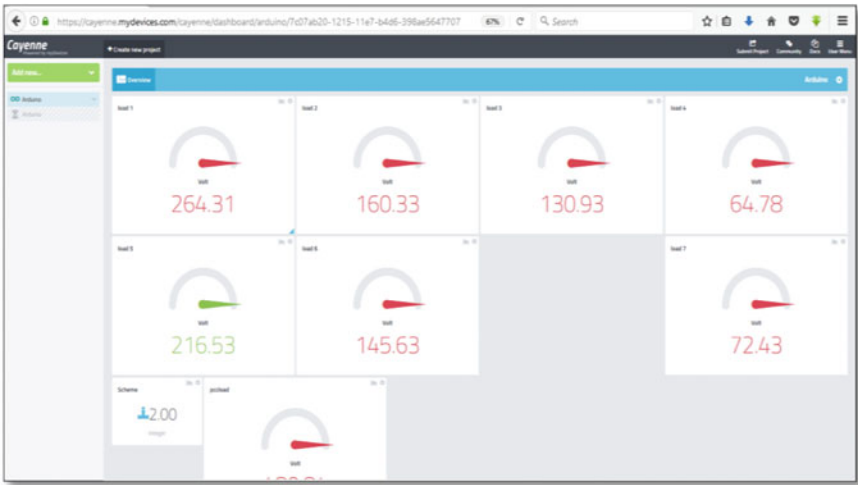


Fig. 12.4 Remote monitoring station

The sensing pin and monitoring/control pin are connected appropriately. The Ethernet shield module is mounted on the Arduino board with supply and ground connections. The pin connections for the Ethernet shield module are done. The test system parameters are proportionally scaled down to a 5 V prototype network. In case at a node, the load current begins to rise and crosses the permissible limit, then the fault condition is detected. In order to clear the fault, the MPC sends a LOW (logic 0) signal through the control pin to the corresponding relay. This signal de-energizes the coil and the relay acts like an open switch and disconnects the load from the system. The system status is updated in the MPC. Once the device is connected to the Cayenne platform, the remote monitoring dashboard is designed as in Fig. 12.4.

12.4 Results and Discussion

Remote monitoring system displays the load parameters. Figure 12.5 shows the system status and load characteristics as the load is displayed on the Arduino IDE. When an overcurrent fault is detected at load 6, a fault-clearing sequence is initiated at the MPC and it disconnects the faulted node from the distribution grid. Figure 12.6 shows the system status after fault is cleared, disconnecting CB6 and thus loads 6 and 7 are inactive.

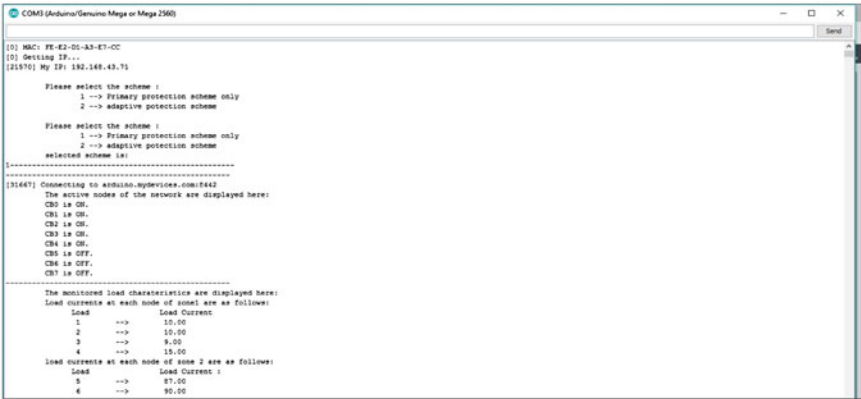


Fig. 12.5 MPC load display

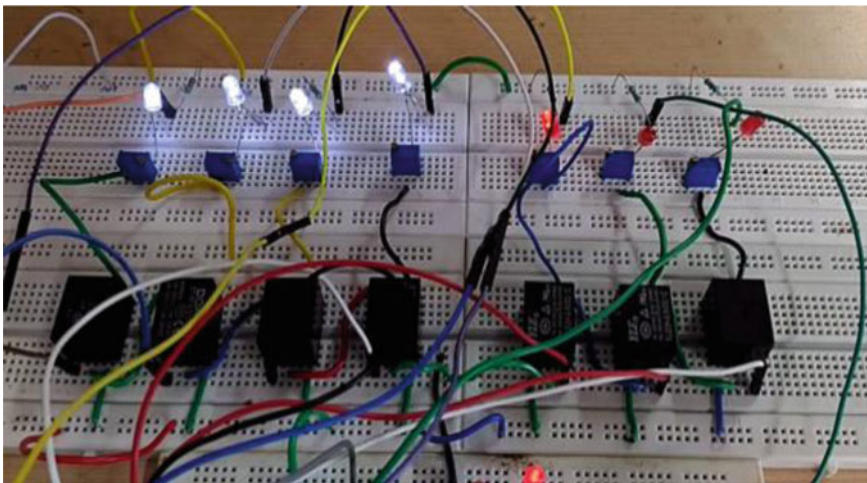


Fig. 12.6 Fault clearance for fault at load 6

12.5 Conclusion

In this chapter a hardware prototype of the test was realized, tested and validated with adaptive protection scheme. The microgrid protection centre was created using Arduino ATmega 2560, which was capable of system monitoring, identifying fault location and measuring the fault current value. The status of the system was then sent to a remote device using IOT. The remote device displayed the real-time values of current (both normal and fault values), using the open source platform, Cayenne. The proposed adaptive protection scheme may be extended to larger networks conveniently.

References

1. Swathika OG, Hemamalini S (2016) Prims aided Dijkstra algorithm for adaptive protection in microgrids. *IEEE J Emerg Sel Top Power Electr* 4(4):1279–1286
2. Arya R, Yadav R, Agarwal R, Swathika OG (2018) Dijkstra's algorithm for shortest path identification in reconfigurable microgrid. *J Eng Appl Sci* 13(3):717–720
3. Mahapatra S, Swathika OG (2018) Hybrid prims-Viterbi's algorithm for protecting multiple utility grids interfaced microgrid. *J Telecommun Electr Comput Eng* 10(1–9):109–113
4. Saluja J, Biswas S, Roy S, Swathika OG (2018) Performance analysis of graph algorithms for microgrid protection. *J Telecommun Electr Comput Eng* 10(1–8):115–118

Chapter 13

Performance Comparison Between Sensor and Sensorless Control of Permanent Magnet Synchronous Motor with Wide Speed Range of Operations



N. Krishna Kumari, D. Ravi Kumar, and K. Renu

Abstract Knowledge of rotor speed and rotor position is essential for effective functioning of Field-Oriented Control (FOC) technique. But this requires sensors, and thus employing advanced vector control strategies is challenging in terms of cost and reliability. In this chapter, a Sensorless Field-Oriented Control of Permanent Magnet Synchronous Motor (PMSM) covering wide speed ranges is evaluated. Maximum Torque Per Ampere (MTPA) can enhance the torque output capability, minimize the stator current and thereby copper loss and increase the overall drive efficiency. However, it is not suitable for above base speeds due to limitations on inverter ratings. Domestic appliances such as washing machine require higher speeds during spin dry cycle and usually two to three times of rated speeds. Even traction applications require a wide range of speed control. To achieve speeds greater than rated speeds, Flux Weakening is employed. Model Reference Adaptive System (MRAS) based on stator current controller is used as rotor position estimation algorithm and a comparison is performed between a controller with sensors and a sensorless controller. Results have demonstrated the effectiveness of the sensorless PMSM compared to controller with sensors for both below and above base speed operations.

Keywords Permanent Magnet Synchronous Motor (PMSM) · Field-Oriented Control (FOC) · Model Reference Adaptive System (MRAS) · Maximum Torque Per Ampere (MTPA)

N. K. Kumari · D. R. Kumar (✉) · K. Renu
VNR Vignana Jyothi Institute of Engineering and Technology, Hyderabad, India
e-mail: krishnakumari_n@vnrvjiet.in; ravikumar_d@vnrvjiet.in

© Springer Nature Switzerland AG 2020
L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_13

Abbreviations

FOC	Field-Oriented Control
MRAS	Model Reference Adaptive System
MTPA	Maximum Torque Per Ampere
PMSM	Permanent Magnet Synchronous Motor

13.1 Introduction

Permanent Magnet Synchronous Motor (PMSM) is gaining more importance mainly due to its attractive features such as high power density, high efficiency, high torque to weight ratio and so on [1–3]. PMSM is one of the fastest-growing members of the variable speed drive family and is extensively studied among researchers, scientists and engineers [3]. Maintenance-free operation, robustness against environment, compact size and high controllability are some more features of PMSM that are responsible for its wide utilization in traction applications, in electric vehicles and hybrid EVs [2, 4, 5], in domestic appliances like washing machine etc.

Field-Oriented Control (FOC) technique provides an independent control of torque and flux components in a manner similar to a DC motor. However, knowledge of rotor position and speed information is essential for its effective functioning. This requires sensors that increase the cost, size and weight of complete drive system [4, 6] and reduce the reliability [7].

Knowledge of the rotor flux position is the core of the FOC for its proper functioning [8, 9]. Applications like washing machine need to run at two to three times the rated speeds during spinning cycles and are subjected to mechanical vibration and electric power oscillation, which result in control complexity [10]. Similarly, in case of electric vehicles, there can be mechanical wear out of sensors due to harsh/muddy environment [11], it gets exposed to and this results in incorrect rotor position. However, usage of high-resolution optical encoders or electromagnetic resolvers results in an increased cost of the overall drive system and reduced reliability especially in harsh environments.

Sensorless techniques are classified into two broad categories: [9].

- (i) Based on motor model.
- (ii) Based on rotor saliency information.

In this work, Model Reference Adaptive System (MRAS) based on stator current controller is chosen as the speed and position estimation algorithm due to its simplicity in implementation and accuracy in the estimated results.

In this work, it is aimed to achieve a wide speed range of operation (below and above base speeds) while minimizing the torque ripples and providing high dynamic performance with FOC of PMSM. Sensorless control algorithms are adopted to streamline the development of variable-speed PMSM controls for inexpensive consumer applications and to provide reliable operation.

This chapter is organized as follows. Section 13.2 gives explanation on MRAS technique for both Maximum Torque Per Ampere (MTPA) and flux-weakening operations. Section 13.3 focuses on analysis of results obtained with qualitative analysis of the speed and torque responses of PMSM and provides a comparative analysis between sensor and sensorless controllers. Section 13.4 gives conclusions drawn from this work along with future work and recommendations. The work is implemented in MATLAB software environment.

13.2 MRAS Using MTPA and Flux Weakening Operations

In this work, MRAS is used as a sensorless technique to estimate the speed of a PMSM. It consists of two models, namely reference model and adjustable model. Reference model is independent on the estimated variable, whereas the adjustable model depends on the estimated variable. The adaptation mechanism generates the difference between the two signals to tune the estimated variable and sends to the adjustable model. The estimated value in this way is driven to its true value [12]. The block diagram of MRAS is as shown in Fig. 13.1.

Below base speeds, MTPA is suitable for PMSM drive since it reduces copper losses [5] and increases the efficiency. Inverter DC bus voltage will be sufficient enough to counteract the back EMF of PMSM. In PMSM, field flux cannot be reduced directly as it is created by permanent magnets. Hence the magnetic flux perceived by stator windings is reduced by imposing negative d -axis current. This weakens the net air gap flux and is referred to as flux weakening. The negative d -axis current component for weakening the flux is given in Eq. (13.1).

$$i_{dref} = -\frac{\lambda_f}{L_d} + \sqrt{\left(\frac{V_{dc}}{\omega_r}\right)^2 - (L_q i_{qref})^2} \tag{13.1}$$

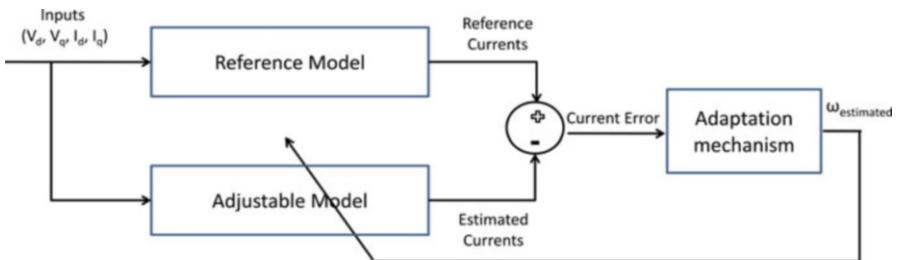


Fig. 13.1 Illustration of the MRAS

When using MRAS to estimate the rotor speed and position of a PMSM, a motor variable, for example, stator current or power, should be chosen as the state variable [13]. In this work, stator currents are chosen as state variables. Stator current model of surface PMSM in DQ or rotor reference frame [14] is given in Eqs. (13.2, 13.3, 13.4 and 13.5).

$$V_d = R_s i_d + L_s \frac{d}{dt} i_d - \omega_e L_s i_q \quad (13.2)$$

$$V_q = R_s i_q + L_s \frac{d}{dt} i_q + \omega_e L_s i_d + \omega_e \lambda_f \quad (13.3)$$

On Simplifying Eq. (13.2) and Eq. (13.3)

$$\frac{di_d}{dt} = -\frac{R_s}{L_s} i_d + \omega_e i_q - \frac{V_d}{L_s} \quad (13.4)$$

$$\frac{di_q}{dt} = -\frac{R_s}{L_s} i_q - \omega_e i_d - \frac{\omega_e \lambda_f}{L_s} - \frac{V_q}{L_s} \quad (13.5)$$

Where V_d , V_q are stator voltage components in dq frame of axes i_d , i_q are stator current components; L_s is stator inductance; R_s is stator resistance; λ_f is rotor flux; ω_e is rotor speed.

The following matrix Eq. (13.6) is obtained from Eqs. (13.4) and (13.5):

$$\frac{d}{dt} \begin{bmatrix} i_d + \frac{\lambda_f}{L_s} \\ i_q \end{bmatrix} = \begin{bmatrix} -\frac{R_s}{L_s} & \omega_e \\ -\omega_e & -\frac{R_s}{L_s} \end{bmatrix} \begin{bmatrix} i_d + \frac{\lambda_f}{L_s} \\ i_q \end{bmatrix} + \frac{1}{L_s} \begin{bmatrix} V_d + \frac{R_s \lambda_f}{L_s} \\ V_q \end{bmatrix} \quad (13.6)$$

Assuming

$$\left. \begin{aligned} i_d' &= i_d + \frac{\lambda_f}{L_s} \\ i_q' &= i_q \\ V_d' &= V_d + \frac{R_s \lambda_f}{L_s} \\ V_q' &= V_q \end{aligned} \right\} \quad (13.7)$$

Equation (13.6) can be rewritten as

$$\frac{d}{dt} \begin{bmatrix} i_d' \\ i_q' \end{bmatrix} = \begin{bmatrix} -\frac{R_s}{L_s} & \omega_e \\ -\omega_e & -\frac{R_s}{L_s} \end{bmatrix} \begin{bmatrix} i_d' \\ i_q' \end{bmatrix} + \frac{1}{L_s} \begin{bmatrix} V_d' \\ V_q' \end{bmatrix} \quad (13.8)$$

In short, the matrix form in Eq. (13.8) can be written as

$$p i' = A i' + B u' \quad (13.9)$$

Speed information is included in state matrix A, which is to be identified. PMSM itself is chosen as the reference model. Adjustable model is used to estimate the stator currents using Eq. (13.10).

$$\frac{d}{dt} \begin{bmatrix} \hat{i}_d \\ \hat{i}_q \end{bmatrix} = \begin{bmatrix} -\frac{R_s}{L_s} & \hat{\omega}_e \\ -\hat{\omega}_e & -\frac{R_s}{L_s} \end{bmatrix} \begin{bmatrix} \hat{i}_d \\ \hat{i}_q \end{bmatrix} + \frac{1}{L_s} \begin{bmatrix} \hat{V}_d \\ \hat{V}_q \end{bmatrix} \quad (13.10)$$

In short, the above Eq. (13.10) can be written as

$$p \hat{i} = A \hat{i} + B u \quad (13.11)$$

The adaptive mechanism is the commonly used PI-controller. This is because it is simple to implement while still performing good, it ensures the stability of the estimator.

The generalized error is given by $e = \hat{i} - i'$.

By analysing this error together with the Popov hyperstability theorem, speed can be estimated using the following adaptive law:

$$\hat{\omega}_e = \left(K_{pMRAS} + \frac{K_{iMRAS}}{s} \right) \left(i_d' \hat{i}_q - i_q' \hat{i}_d \right) \quad (13.12)$$

By substituting Eq. (13.7) in Eq. (13.12), the final estimated speed is obtained as in Eq. (13.13).

$$\hat{\omega}_e = \left(K_{pMRAS} + \frac{K_{iMRAS}}{s} \right) \left(i_d \hat{i}_q - i_q \hat{i}_d - \frac{\lambda_f}{L_s} (i_q - \hat{i}_q) \right) \quad (13.13)$$

Block diagram of the proposed work using MRAS as rotor position estimation algorithm is shown in Fig. 13.2. Once the speed is estimated, rotor position is acquired by integrating it and is given in Eq. (13.13).

$$\hat{\theta} = \int \hat{\omega}_e dt \quad (13.14)$$

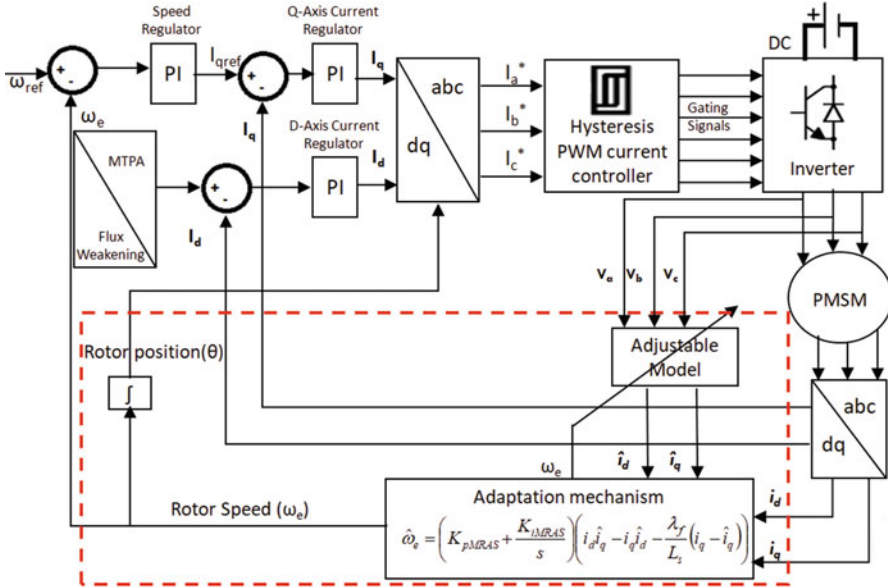


Fig. 13.2 Block diagram of the proposed work

Table 13.1 Parameters of a PMSM

Parameters	Values
Power rating	2 kW
Stator resistance (R_s)	2.8750 Ω
Inductance d -axis (L_d)	0.0085 H
Inductance q -axis (L_q)	0.0085 H
DC voltage (V_{dc})	160 V
Rotor flux (λ)	0.175 Wb
Moment of inertia (J)	0.0008 kg/m ²
Friction (B)	0.001 nm/rad
Poles (P)	4
Load torque (T_L)	6 N-m

13.3 Results and Discussion

The parameters of a PMSM are given in Table 13.1. The main objective of this work is to obtain the following:

- (i) Better dynamic response of the PMSM drive using FOC.
- (ii) Provide a wide range of speed control with flux-weakening strategy.
- (iii) Validate the results with and without sensors under various operating conditions.
- (iv) Minimize the torque and flux ripples for various dynamic conditions.

Response of stator phase currents, its dq currents, torque, speed and flux waveforms are plotted for MTPA and flux weakening and shown in Fig. 13.3. Similar

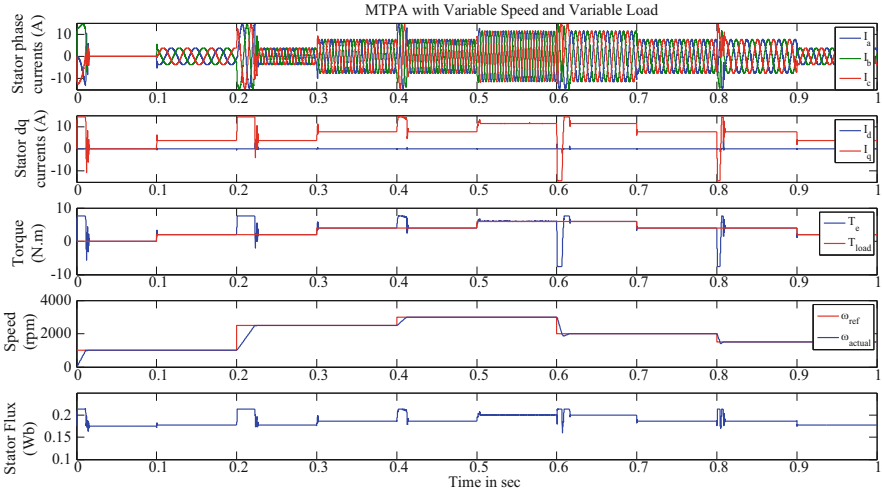


Fig. 13.3 Response of stator currents, direct current, quadrature current, speed, torque and flux waveforms under MTPA control with variable load and variable speed with sensors

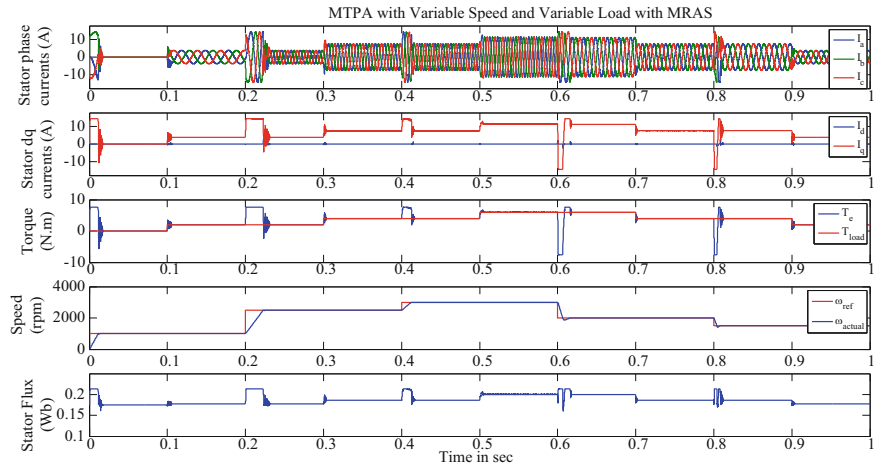


Fig. 13.4 Response of stator currents, direct current, quadrature current, speed, torque and flux under MTPA control with variable load and variable speed with sensorless control

conditions are applied by using MRAS technique and responses are plotted in Fig. 13.4. Torque ripple and flux ripple percentages are calculated and are tabulated in Table 13.2 for sensorless mode of operation.

Table 13.2 PMSM drive performance at variable load and speed for below base speeds using MTPA

Time (s)	Speed (rpm)	Load (N.m)	I_s (A)		I_d (A)		I_q (A)		Torque ripple (%)		Flux ripple (%)	
			With Sensors	Without sensors	With sensors	Without sensors	With sensors	Without sensors	With sensors	Without sensors	With sensors	Without Sensors
0.1	1000	2	3.8	3.8	0	0	3.8	3.8	0.7	0.65	0.06	0.06
0.2	2500	2	3.8	3.8			3.8	3.8	1.28	1.2	0.11	0.11
0.3		4	7.6	7.6			7.6	7.6	0.73	0.68	0.11	0.11
0.4	3000	4	7.6	7.6			7.6	7.6	0.88	0.83	0.21	0.16
0.5		6	11.4	11.4			11.4	11.4	1.5	1.66	0.45	0.45
0.6	2000	6	11.4	11.4			11.4	11.4	0.43	0.42	0.15	0.15
0.7		4	7.6	7.6			7.6	7.6	0.63	0.55	0.11	0.11
0.8	1500	4	7.6	7.6			7.6	7.6	0.48	0.49	0.05	0.05
0.9		2	3.8	3.8			3.8	3.8	0.84	0.82	0.06	0.11

13.3.1 MTPA with Variable Speed and Variable Load

- (i) Dynamic behaviour of PMSM drive is shown in Fig. 13.5. From Fig. 13.5, it can be noticed that near sinusoidal currents are being drawn.
- (ii) It is also observed that there is a heavy in-rush of currents at the beginning and they settle at some constant magnitudes once the commanded speed is met.
- (iii) They remain at near zero magnitudes as there is no load applied till 0.1 s.
- (iv) As and when torque is increased or decreased, stator current magnitudes are changing proportionally so as to meet the commanded torque.
- (v) Similarly, as and when speed is increasing, large current magnitudes are applied such that the response time for reaching the commanded speed is reduced. They settle at some constant magnitudes once commanded speed is obtained.
- (vi) Frequency of operation is maintained constant for a given constant speed duration. As the speed increases, torque ripple as well as flux ripple is increased.
- (vii) As load torque is increased, torque ripple is reduced whereas flux ripple is increased. There is not much difference being seen in torque ripple and flux ripple for sensor and sensorless control.

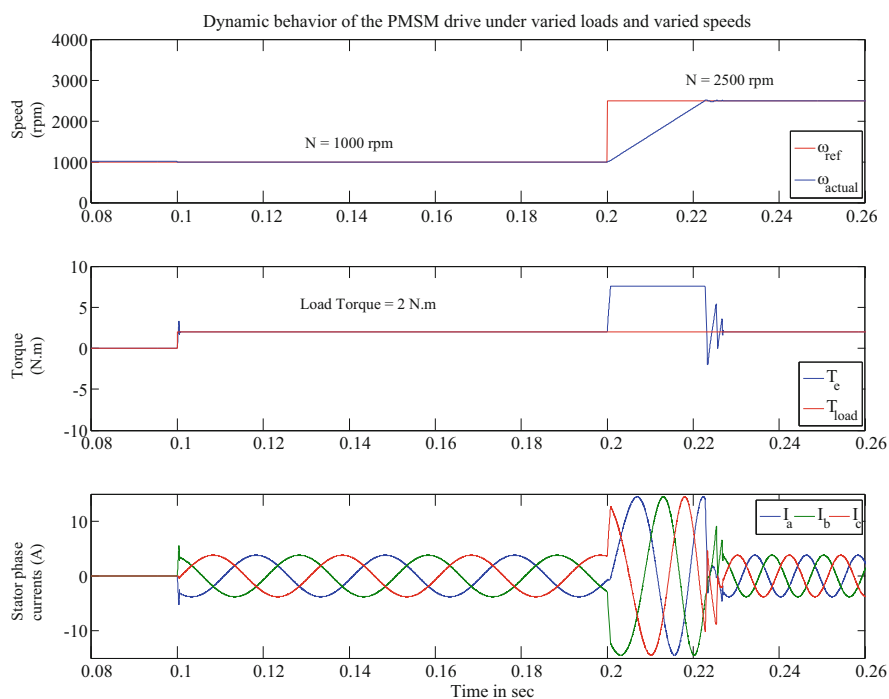


Fig. 13.5 Dynamic behaviour of PMSM drive between 0.08 and 0.26 s for variable loads and variable speeds

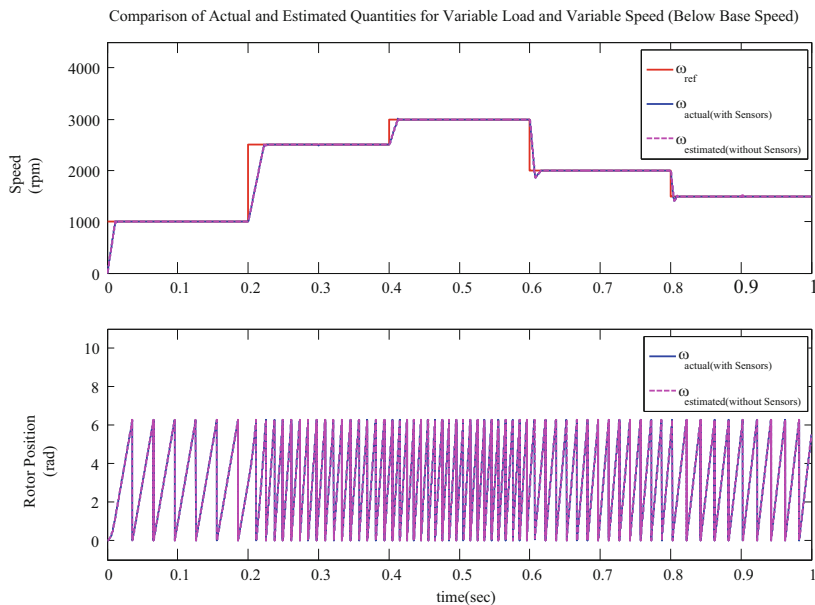


Fig. 13.6 Comparison of actual and estimated speed and position quantities under variable loads and variable speeds

(viii) Figure 13.6 shows a comparison between actual and estimated speed quantities. From Fig. 13.6, it can be observed that speed and rotor position quantities that are estimated using MRAS technique are exactly tracing the actual/measured quantities and there is no error that can be seen.

13.3.2 Flux Weakening with Variable Speed and Variable Load

The dynamic performance of the drive is tested considering variable load torque and speeds. Figures 13.7 and 13.8 shows the responses of stator current, direct axis current, quadrature axis current, electromagnetic torque, speed and flux waveforms for both sensor and sensorless modes respectively.

Flux weakening principle is applied in this case since all the commanded speeds are above the base speed. This will weaken the net flux perceived by stator windings and helps in achieving speeds above base speeds. Figure 13.9 shows a comparison between actual and estimated speed and position quantities. Torque ripple and flux ripple percentages are calculated in manner similar to Sect. 13.3.1 and are tabulated in Table 13.3 for sensorless modes of operation. The following observations are made.

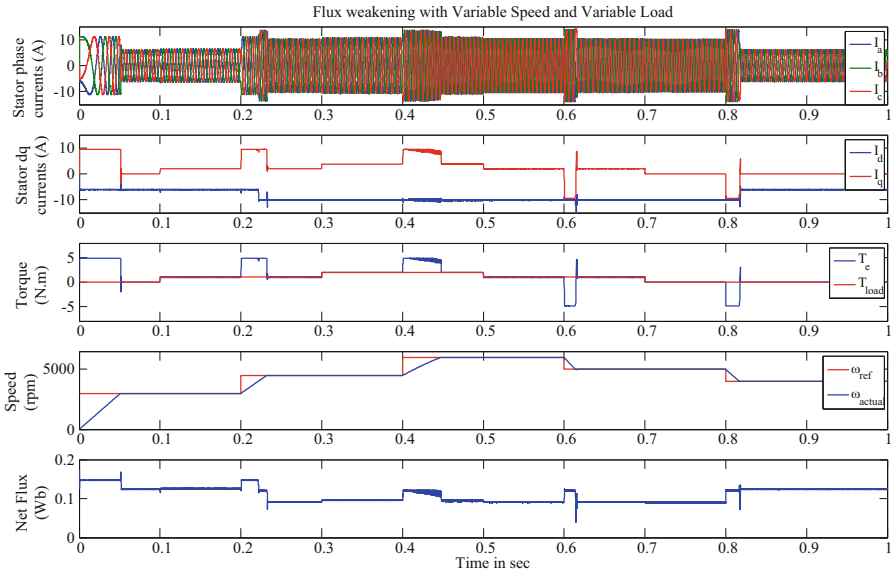


Fig. 13.7 Response of stator currents, direct current, quadrature current, speed, torque and flux for flux weakening with variable load and variable speed with sensors

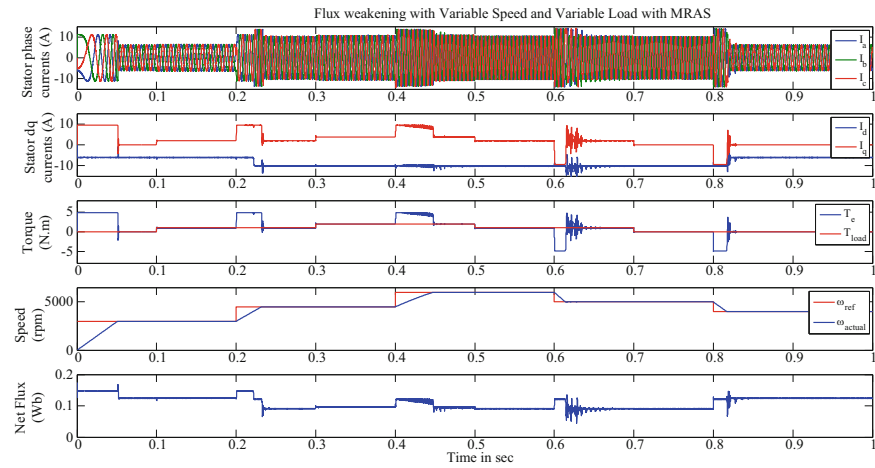


Fig. 13.8 Response of stator currents, direct current, quadrature current, speed, torque and flux for flux weakening with variable load and variable speed with sensorless control

- (i) When the motor speed is increased beyond base speeds, that is, 4500 rpm at $t = 0.2$ s, 6000 rpm at $t = 0.4$ s; the d -axis component of stator current is moving towards negative axis, which can be observed in the waveform as the value of i_d is reduced and thereby flux weakening can be viewed properly by observing flux waveform.

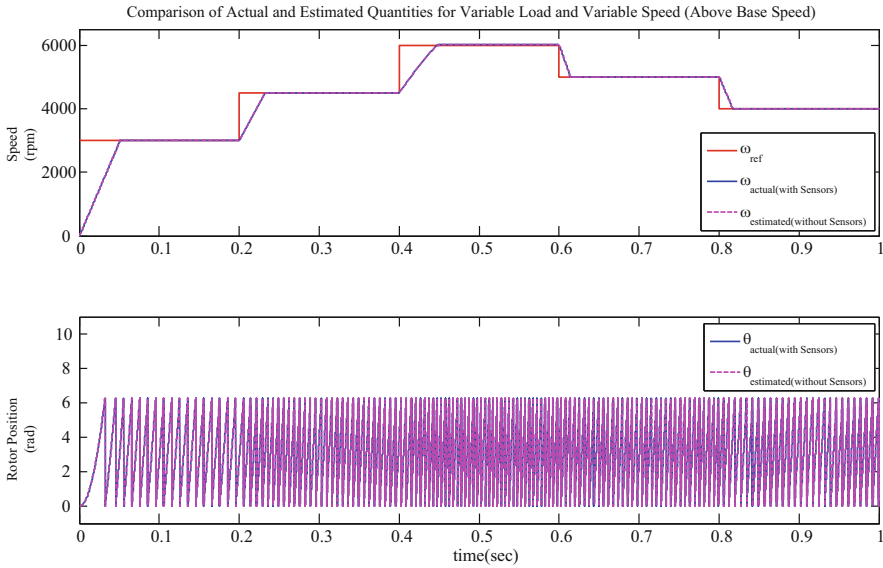


Fig. 13.9 Comparison of actual and estimated speed and rotor position quantities for flux weakening control for variable load and variable speeds

- (ii) But, as the load torque is varying at above specified time intervals, correspondingly only q -axis component of stator current is varying keeping d -axis unaffected within that speed range. Thus, it is verified that, to drive load torques at beyond rated speeds, the value of i_d must be changed according to desired speed.
- (iii) Torque ripple and flux ripple percentages are considerably high for above base speeds when compared to ripples at below base speeds.
- (iv) Ripple percentages are slightly high for MRAS control when compared to without MRAS control.
- (v) From Fig. 13.9, it can be observed that speed and position quantities that are estimated using MRAS technique are exactly tracing the actual/measured quantities and there is no error that can be seen.

13.4 Conclusion

Rotor speed and position are estimated using MRAS technique in this work. Different operating conditions are applied at below and above rated speeds and the drive performance is validated in sensor and sensorless modes. From the waveform quality and torque ripple percentages in Tables 13.2 and 13.3 for both below and above base speeds, it is noticed that a similar performance is observed in both sensor and sensorless modes. Results of the proposed sensorless control system have

Table 13.3 PMSM drive performance at variable load and speed for above base speeds using flux weakening

Time (s)	Speed (rpm)	Load (N·m)	$I_s(A)$		$I_d(A)$		$I_q(A)$		Torque ripple (%)		Flux ripple (%)	
			With sensors	Without sensors	With sensors	Without sensors	With sensors	Without sensors	With sensors	Without sensors	With sensors	Without sensors
0.2	4500	1	10.3	10.3	-10.1	-10.1	1.9	1.9	3.37	6.06	2.52	2.85
0.3		2	10.75	10.75			3.8	3.8	2.06	2.11	2.41	2.20
0.4	6000	2	10.8	10.8	-10.3	-10.3	3.8	3.8	6.09	13.93	7.28	6.93
0.5		1	10.3	10.3			1.9	1.9	6.07	6.58	4.88	4.08
0.6	5000	1	10.3	10.3	-10.1	-10.1	1.9	1.9	3.93	9.76	2.63	8.56

demonstrated its effectiveness and fast dynamic response. It can be concluded that FOC in sensorless mode appears to be more viable approach since it enables the minimization of cost barrier, weight, size and complexity and increases reliability and robustness of the drive. Torque ripple percentages are slightly high in flux weakening region and this work can be extended to reduce the ripple by adopting hybrid fuzzy PI controller.

References

1. Halder S, Agarwal P, Srivastava SP (2015) Comparative analysis of MTPA and ZDAC control in PMSM drive. Annual IEEE India conference (INDICON), pp 17–20
2. Li Z, Li H (2012) MTPA control of PMSM system considering saturation and cross-coupling. 15th international conference on electrical machines and systems (ICEMS), Sapporo, pp 1–5
3. Tahami F, Nademi H, Rezaei M (2008) A high-performance vector-controlled PMSM drive with maximum torque per ampere operation. IEEE 2nd international power and energy conference, Johor Bahru, pp 254–258
4. Lidozzi A, Solero L, Crescimbin F, Napoli AD (2007) SVM PMSM drive with low resolution hall-effect sensors. IEEE Trans Power Electron 22(1):282–290
5. Wang J, Wu J, Gan C, Sun Q (2016) Comparative study of flux-weakening control methods for PMSM drive over wide speed range. 19th international conference on electrical machines and systems (ICEMS), Chiba, Japan, pp 1–6
6. Nak H, Gülbahce MO, Gokasan M, Ergene AF (2015) Performance investigation of extended Kalman filter based observer for PMSM using in washing machine applications. 9th International conference on electrical and electronics engineering (ELECO), Bursa, pp 618–623
7. Karakas S, Gokasan M (2011) Online fine tuning of current PI controllers in spin-dry cycle of washing machines. International Aegean conference on electrical machines and power electronics and electro motion, Joint conference, Istanbul, pp 742–747
8. Sandeep SA, Prabhu LV, Renukappa NM, Suryanarayana K (2016) Speed control of IPMSM motor without rotor position sensors based on MRAS observer with space vector modulation. IEEE international conference on recent trends in electronics, information & communication technology (RTEICT), Bangalore, pp 1841–1845
9. Rashed M, MacConnel PFA, Stronach AF, Acarnley P (2007) Sensorless indirect-rotor-field-orientation speed control of a permanent-magnet synchronous motor with Stator-resistance estimation. IEEE Trans Ind Electron 54(3):1664–1675
10. Chi S, Zhang Z, Xu L (2009) Sliding-mode sensorless control of direct-drive PM synchronous motors for washing machine applications. IEEE Trans Ind Appl 45(2):582–590
11. Hassan M, Mahgoub O, El Shafei A (2013) ANFIS based MRAS speed estimator for sensorless control of PMSM. Brazilian power electronics conference, Gramado, pp 828–835
12. Li Y, Hao Z (2008) Sensorless control of permanent magnet synchronous motor – a survey. IEEE vehicle power and propulsion conference, Harbin, pp 1–8
13. Aygun H, Gokdag M, Aktas M, Cernat M (2014) A novel sensorless field oriented controller for permanent magnet synchronous motors. IEEE international symposium on industrial electronics, pp 715–720
14. Wan J, Li G, Yuan C, Shen H (2009) MRAS speed identification for PMSM based on fuzzy PI control. IEEE conference on industrial electronics and applications, pp 1995–1998

Chapter 14

Passive Fault-Tolerant Control Based on Interval Type-2 Fuzzy Controller for Coupled Tank System



Himanshukumar R. Patel and Vipul A. Shah

Abstract In this chapter, a robust controller for a coupled tank-level control is proposed in presence of system component fault. For this purpose, interval type-2 fuzzy logic control approach (IT2FLC) technique is used to design a controller, named passive Fault Tolerant Interval Type-2 Fuzzy Controller (PFTIT2FLC) based on the robust controller to fault tolerant of coupled tank level control system. The proposed control scheme allows avoiding modelling, reducing the rules number of the fuzzy controller. The simulation results show that the PFIT2FLC can provide good tracking performance, even in presence of system component faults.

Keywords System component fault · Interval type-2 fuzzy logic · Fault-Tolerant Control

Abbreviations

CTLCS	Coupled Tank-Level Control System
FTC	Fault-Tolerant Control
IT2FLC	Interval Type-2 Fuzzy Logic-Control Approach
KMA	Karnik–Mendel Algorithm
LMPC	Linear Model Predictive Control
LQR	Linear Quadratic Regulator
PFTIT2FLC	Passive Fault-Tolerant Interval Type-2 Fuzzy Controller
PID	Proportional Integral Derivative
T1FLC	Type-1 Fuzzy Logic Controller

H. R. Patel (✉) · V. A. Shah
Dharmsinh Desai University, Nadiad, Gujarat, India
e-mail: himanshupatel.ic@ddu.ac.in

14.1 Introduction

The actuator and system component faults in any control system may degrade performance drastically even creating dangerous situation. To tolerate such kind of unwanted situation, Fault-Tolerant Control (FTC) strategy is used to maintain system stability and control performance at an acceptable level. The FTC is classified into two broad categories based on the working principle. One is active FTC and second one is passive FTC [1]. The active FTC required separate algorithm for detection, identification and diagnosis of the fault, and based on the outcome from the algorithm, controller will change the parameters [2]. Contrarily the passive FTC is robust controller that was designed on the basis of predetermined faults [3].

Control design for single-tank- or multiple-tank-level control system with interacting and non-interacting configuration has been a topic of active research in recent years due to their important applications. The coupled-tank-level control system prototype is often used in chemical processing industries and education for the design and implementation of control algorithms. In our study we consider the coupled-tank-level control process laboratory set-up. Because this coupled-tank has non-linear dynamics as well as interaction between two tanks, the control of this system is a challenging task. Many researchers have been interested in the control of coupled-tank level system. Some of them designed trajectory-tracking control strategy for double-tank-level process based on predictive observer [4]. The backstepping controller is designed for coupled-tank-level control system using an adaptive high gain observer in [5], and experimental validation investigates on coupled-tank system in [6]. Also, linear model predictive control (LMPC) strategy is proposed for non-linear coupled-tank-level control process in [7].

Since the last two decades, fuzzy controllers are used more and more for the controlling of coupled-tank-level system. However, it is not easy to understand the organization of the fuzzy rule base, since the fuzzy rules are more complicated than the rules based on the common sense. Moreover, the number of rules is large and the complexity of the fuzzy controllers is high. In [8], authors design type-1 fuzzy logic control for coupled tank level control system, and in [9], type-2 fuzzy logic control is proposed for level control system. To accommodate the different faults like system component and actuator fault, the authors [10] designed PFTC using fuzzy logic and conventional PI controller and reported good control performance in simulation platform with system fault and unknown process disturbance. Thereafter the authors [11] validate the proposed PFTC strategy implemented on real-time single-tank non-interactive system with system fault.

This chapter attributes passive FTC using Interval Type-2 Fuzzy Logic Control (IT2FLC) for Coupled-tank-Level Control System (CTLCS) subject to system component (leak fault) fault.

The rest of the chapter is organized as follows. The dynamics of the coupled-tank-level process is described in Sect. 14.2. The background of the type-2 fuzzy logic control and the design of Passive Fault-Tolerant Interval Type-2 Fuzzy Logic Control (PFTIT2FLC) are presented in Sect. 14.3. The simulation results

demonstrating the effectiveness of the proposed approach are presented in Sect. 14.4. A conclusion of this work is given in Sect. 14.5.

14.2 Model Description of the Coupled-Tank System

14.2.1 Coupled-Tank-Level Control System

As depicted in Fig. 14.1 below, the CTLCS model consists of two cylindrical tanks with the same transversal area and height. The tanks are interconnected through a cylindrical pipe and have individual constant output flows. They are also equipped with two-level sensors. All the flow pipes are equipped with manually adjustable valves. The design objective is to maintain the liquid level of tank 2, (h_2), at a desired reference by controlling the input flow of tank 1, (q_1).

The CTLCS system considered with one fault is the system component (tank 2 bottom leak) fault

14.2.2 CTLCS Mathematical Modelling

The modelling of the CTLCS is carried out by the mass-balance equation and Bernoulli’s equation. At any instance, the rate of change of volume of liquid present inside the tank can be expressed in terms of the liquid inlet, liquid outlet and the tank interaction [2]. Mathematically,

$$A_1 \frac{dh_1}{dt} = q_1 - q_{01} - q_{12} \& A_2 \frac{dh_2}{dt} = q_{12} - q_2 \tag{14.1}$$

According to Bernoulli’s equation, the flow rates [2].

Fig. 14.1 Coupled-tank level control system prototype subject to one fault

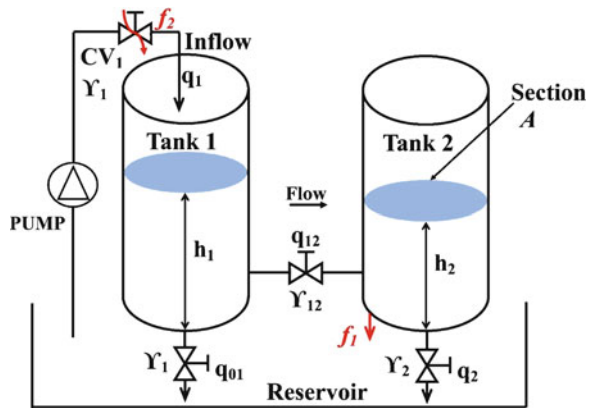


Table 14.1 Coupled-tank-level control system parameters.

A_1 and A_2	Area of tank 1 and tank 2	0.0270 m ²
q_1	Inlet flow rate	0.000162 m ³ /s
Y_1	Discharge coefficient of tank 1	6.3795
Y_2	Discharge coefficient of tank 2	1.614
Y_{12}	Discharge coefficient of coupling valve	4.372
h_1	Operating point of tank 1	0.35 m
h_2	Operating point of tank 2	0.31 m

q_{01} , q_{12} and q_2 are given by

$$q_{01} = Y_1 \sqrt{h_1}, q_2 = Y_2 \sqrt{h_2}, \& q_{12} = Y_{12} \sqrt{h_1 - h_2} \quad (14.2)$$

The CTLCS operating parameters and system parameters are presented in Table 14.1.

The CTLCS model is linearized around the operating point using Taylor's series expansion method. The simulation is carried out with and without faults for regulatory and reference trajectory-tracking control. The transfer function of CTLCS is found on the basis of operating parameters as given in Table 14.1, which is given by (14.3)

$$G_P(\text{CTLCS}) = \frac{2.3721}{S^2 + 2.4915S + 12.8371} \quad (14.3)$$

14.3 PFTIT2FLC design and Background of Type-2 FLC

14.3.1 Background of the Type-2 Fuzzy Logic Control

Type-1 and type-2 fuzzy logics are mainly similar. The only essential difference between them is the membership functions shape, besides the output process. Indeed, an interval type-2 fuzzy controller consists of a fuzzifier, an inference engine, a rules base, a type reduction and a defuzzifier [12–14]. The block diagram of the type-2 FLC is given in [19].

14.3.1.1 Fuzzifier

The fuzzifier maps the crisp input vector $(e_1, e_2, \dots, e_n)^T$ to a type-2 fuzzy system \tilde{A}_x , which is very similar to the procedure performed in a type-1 fuzzy logic system.

Rules.

The general form of the i th rule of the type-2 fuzzy logic system can be written as

$$\text{If } e_1 \text{ is } \tilde{F}_1^i \text{ and } e_2 \text{ is } \tilde{F}_2^i \text{ and } \dots e_n \text{ is } \tilde{F}_n^i, \text{ Then } y^i = \tilde{G}^i \quad i = 1, \dots, M \quad (14.4)$$

where \tilde{F}_j^i represents the type-2 fuzzy system of the input state j of the i th rule, x_1, x_2, \dots, x_n are the inputs, \tilde{G}^i is the output of the type-2 fuzzy system for the rule i , and M is the number of rules.

14.3.1.2 Type Reducer

After definition of the rules and executing the inference, the type-2 fuzzy system resulting in type-1 fuzzy system is computed. In this part, the available methods to compute the centroid of type-2 fuzzy system using the extension principle are discussed [13]. The centroid of type-1 fuzzy system A is given by

$$C_A = \frac{\sum_{i=1}^n z_i w_i}{\sum_{i=1}^n w_i} \quad (14.5)$$

where n represents the number of discretized domain of A , $z_i \in R$ and $w_i \in [0, 1]$.

If each z_i and w_i is replaced by a type-1 fuzzy system (Z_i and W_i), with associated membership functions of $\mu_Z(z_i)$ and $\mu_W(w_i)$, respectively, and by using the extension principle, the generalized centroid for type-2 fuzzy system \tilde{A} can be expressed by

$$GC_{\tilde{A}} = \int_{z_1 \in Z_1} \dots \int_{z_n \in Z_n} \int_{w_1 \in W_1} \dots \int_{w_n \in W_n} [T_{i=1}^n \mu_Z(z_i) * T_{i=1}^n \mu_W(w_i)] / \frac{\sum_{i=1}^n z_i w_i}{\sum_{i=1}^n w_i} \quad (14.6)$$

14.3.1.3 Defuzzifier

To get a crisp output from a type-1 fuzzy logic system, the type-reduced set must be defuzzified. The most common method to do this is to find the centroid of the type-reduced set. If the type-reduced set Y is discretized to n points, then the following expression gives the centroid of the type-reduced set:

$$y_{\text{output}}(x) = \frac{\sum_{i=1}^n y^i \mu(y^i)}{\sum_{i=1}^m \mu(y^i)} \quad (14.7)$$

The output can be computed using the iterative Karnik Mendel Algorithm (KMA) [12]. Therefore, the defuzzified output of an interval type-2 FLC is

$$y_{\text{output}}(x) = \frac{y_l(x) + y_r(x)}{2} \quad (14.8)$$

14.3.2 PFTIT2FLC Design for Coupled-Tank-Level Control System

In order to eliminate the high oscillation, non-linear system and model uncertainty, a continuous Interval Type-2 Fuzzy logic control (IT2FLC) is used to approximate the discontinued control. The proposed control (PFTIT2FLC) scheme is shown in Fig. 14.2.

The IT2FLC membership functions of the fuzzy input variable are chosen to be triangular for all upper and lower membership functions. The used labels of the fuzzy variable residue, error and its derivative are {very small (VS), small (S), big (B), and very big (VB)}. Figure 14.3 presents the type-2 membership functions for the IT2FLC. The corrective control is decomposed into four levels, the IT2FLC consists of three input and one output and so total rules can be 64. The main motivation to implement the IT2FLC in the proposed system is non-linear uncertain

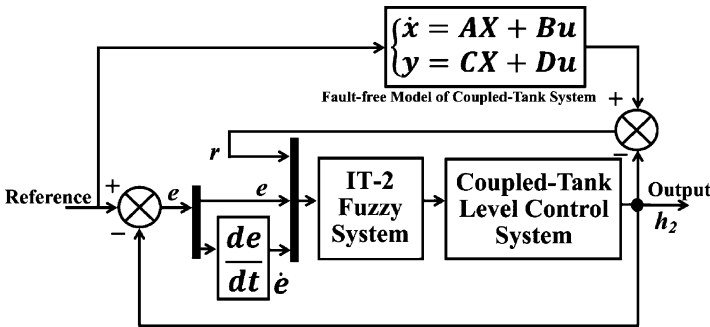


Fig. 14.2 Block diagram of the proposed controller PFTIT2FLC

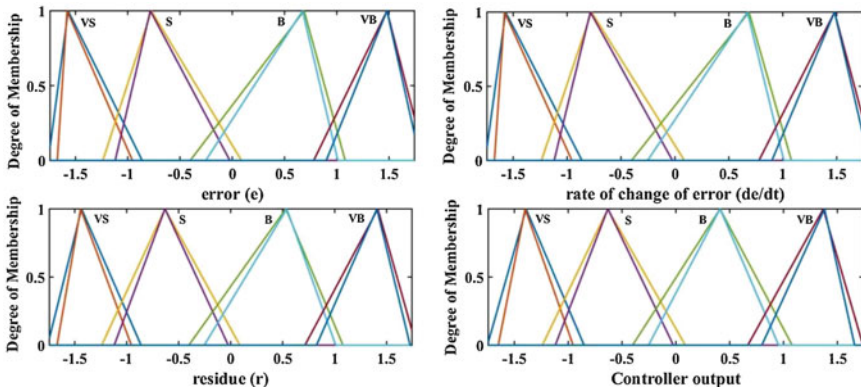


Fig. 14.3 Input–output membership function of IT2FLC

system and does not require accurate system mathematical model. Also IT2FLC itself is fuzzy, with a new dimension called the footprint of uncertainty, and it is best suited for non-linear uncertain system.

14.4 Simulation Results

In order to verify the proposed control scheme for CTLCS subject to system component fault, first fault-free case investigates CTLCS with proposed control scheme and compares with other control scheme. The fault-free response for various controllers is presented in Fig. 14.4 with pulse trajectory, respectively. The pulse trajectory is taken for checking the tracking control abilities for proposed controller and compared with other controller.

14.4.1 Tracking Response with System Component Faults

Two tests examine CTLCS: Test 1 is for pulsed trajectory tracking with system component faults and test 2 is for sine trajectory tracking with system component faults with different magnitudes.

Test 1 responses with various controllers are depicted in Fig. 14.5 with system component fault $f_i = 10$ cm and test 2 responses are depicted in Fig. 14.6 with fault magnitude $f_i = 20$ cm. The comparative error results are also presented in Table 14.2.

To validate the proposed control technique, we considered LQR, PID and T1FLC methods; the LQR and PID techniques are model-based methods where IT2FLC and

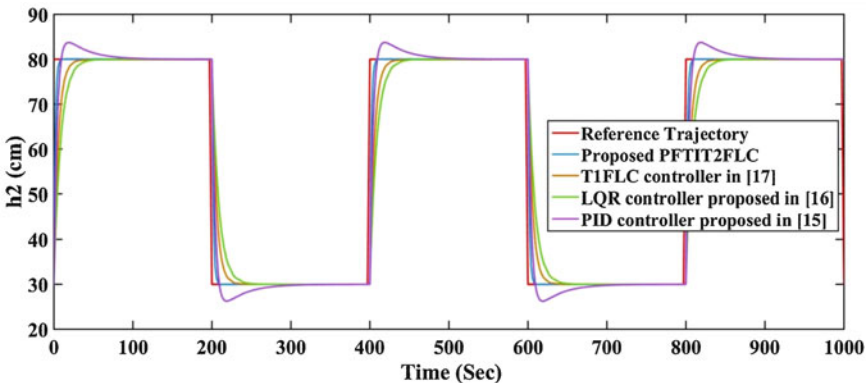


Fig. 14.4 Pulsed trajectory-tracking response comparison without faults

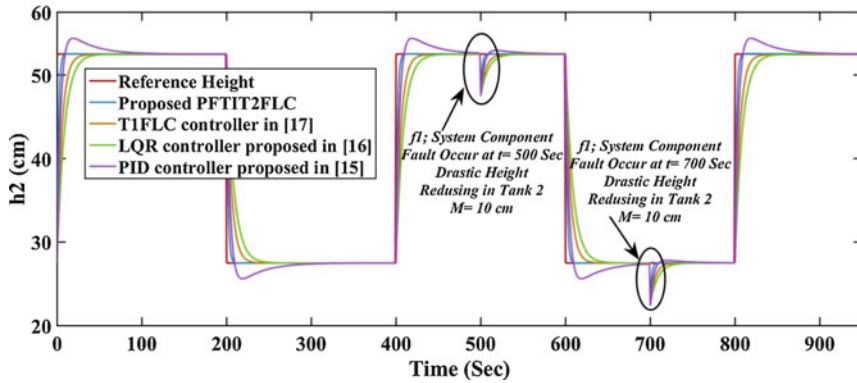


Fig. 14.5 Pulsed trajectory-tracking response comparison with system component faults

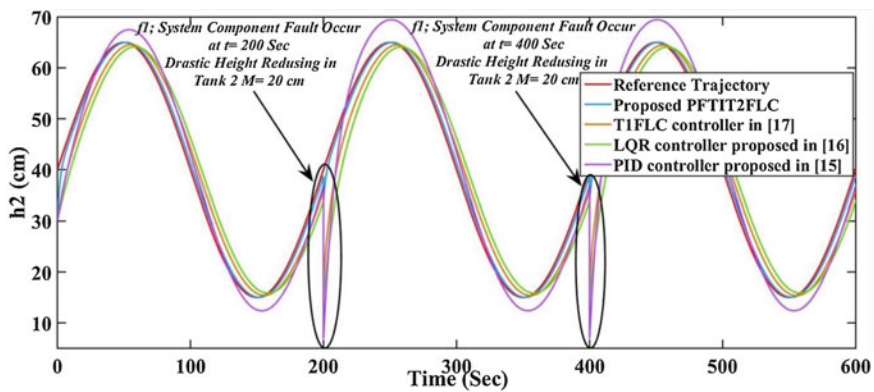


Fig. 14.6 Sine trajectory-tracking response comparison with system component faults

Table 14.2 Quantitative comparison between the proposed controller, PID controller, LQR controller and T1FLC with system component faults.

Test	Control scheme	IAE	ISE
		h_2	h_2
Test 1	PID controller proposed in [15]	5.38	2.303
	LQR control proposed in [16]	3.566	1.579
	T1FLC proposed in [17]	2.31	1.577
	The proposed PFTIT2FLC	1.023	0.472
Test 2	PID controller proposed in [15]	34.52	15.06
	LQR control proposed in [16]	6.091	5.92
	T1FLC proposed in [17]	3.27	2.105
	The proposed PFTIT2FLC	1.672	0.3914

T1FLC techniques are model-free methods. So when problem of accurate mathematical model of the non-linear uncertain system the model free method is the best choice, provided that in FLC proper fuzzy rules are framed.

14.5 Conclusion

In this chapter, performances of Passive Fault-Tolerant interval type-2 fuzzy logic controller (PFTIT2FLC) for coupled-tank-level control system are investigated in the presence of system component (tank leak) fault. Fuzzy controllers are better at representing the real world with abstract fuzzy logic rules that provide the allowance for non-linear unpredictability in systems when fuzzy rules are properly framed and it is model-free techniques, hence proposed IT2FLC gives superior performance as compared to LQR and conventional PID controller. The simulation results have shown high efficiency of this control strategy; it maintains the stability and the good performances of the coupled-tank-level control system in presence of the system component fault. In addition, the comparative study performed with other recent works developed in the literature has shown the effectiveness of the proposed control approach. In the future work the general form of type-2 fuzzy sets proposed in [18, 19] will be introduced in the proposed control to increase robustness of the system and handle the uncertainty and external disturbances.

References

1. Patel HR, Shah VA (2018) Fault detection and diagnosis methods in power generation plants – the Indian power generation sector perspective: an introductory review. *PDP J Energy Manag* 2(2):31–49. ISSN 2581-5849
2. Patel HR, Shah VA (2018) A framework for fault-tolerant control for an interacting and non-interacting level control system using AI. In: *Proceedings of the 15th international conference on informatics in control, automation and robotics – volume 1: ICINCO*, pp 180–190. <https://doi.org/10.5220/0006862001800190>. ISBN:978-989-758-321-6
3. Patel HR, Shah VA (2018) Fault tolerant control systems: a passive approaches for single tank level control system. *I-manager's J Instrum Control Eng* 6(1):11–18. <https://doi.org/10.26634/jic.6.1.13934>
4. Houssemeddine G, Salim Hadj S, Faouzi M (2015) Observer-based predictive liquid level controller for a double tank process. *7th international conference on modelling, identification and control (ICMIC 2015)*, IEEE, 18–20 December 2015, Sousse (Tunisia), pp 1–6
5. Turki A, Hadj Said S, M'Sahli F (2014) Backstepping control for two tanks process based on adaptive high gain observer. *11th international multi-conference on systems signals & devices (SSD)*
6. Pan H, Wong H, Kapila V, de Queiroz MS (2005) Experimental validation of a nonlinear backstepping liquid level controller for a state coupled two tank system. *Control Eng Pract*:27–40

7. Muhammad Usman K, Bilal KM (2012) Liquid level control of nonlinear coupled tanks system using linear model predictive control. International conference on emerging technologies (ICET 2012), pp 1–5. <https://doi.org/10.1109/ICET.2012.6375434>
8. Abid M (2005) Fuzzy logic control of coupled liquid tank system. Information and communication technologies, ICICT 2005. First international conference on, August 2005, pp 144–147
9. Dongrui W, Woei T (2004) A type-2 fuzzy logic controller for the liquid-level process. IEEE international conference on fuzzy systems, IEEE, vol 2, pp 953–958. <https://doi.org/10.1109/FUZZY.2004.1375536>
10. Patel HR, Shah VA (2018) Fuzzy logic based passive fault tolerant control strategy for a single-tank system with system fault and process disturbances. In: Proceedings of the 5th international conference on electrical and electronics engineering (ICEEE), Istanbul, Turkey, May 2018, pp 257–262
11. Patel HR, Shah VA (2019) Performance comparison of passive fault tolerant control strategy with PI and fuzzy control of single-tank level process with sensor and system fault. Science Publications, Am J Eng Appl Sci 12(2):236–246. <https://doi.org/10.3844/ajeassp.2019.236.246>
12. Castillo O, Melin P (2012) A review on the design and optimization of interval type-2 fuzzy controllers. Appl Soft Comput 12:1267–1278
13. Mendel J (2011) Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice-Hall, Springer International Publishing
14. Liang Q, Mendel J (2000) Interval Type-2 fuzzy logic systems: theory and design. IEEE Trans Fuzzy Syst 8(5):535–550
15. Mukherjee D, Kundu P, Gosh A (2016) PID controller design for an interacting tank level process with time delay using MATLABFOMCON toolbox. 2nd international conference on control, instrumentation, energy & communication (CIEC), IEEE, Kolkata, India, pp 1–5. <https://doi.org/10.1109/CIEC.2016.7513803>
16. Dutta S, Seal S, Sengupta A (2014) Real time linear quadratic versus sliding mode liquid level control of a coupled tank system. International conference on devices, circuits and communications (ICDCCom 2014), IEEE, 12–12 September 2014, Ranchi, India, pp 1–6. <https://doi.org/10.1109/ICDCCom.2014.7024741>
17. Bhandare D, Kulkarni NR (2015) Performances evaluation and comparison of PID controller and fuzzy logic controller for process liquid level control. 15th international conference on control, automation and systems (ICCAS 2015) October 13–16, 2015, Busan, South Korea, pp 1347–1352. <https://doi.org/10.1109/ICCAS.2015.7364848>
18. Mehran M, Marziah N (2014) Differentiability of type-2 fuzzy number-valued functions. Commun Nonlinear Sci Numer Simul 19(3):710–725
19. Mehran M, Marziah N (2014) Type-2 fuzzy fractional derivatives. Commun Nonlinear Sci Numer Simul 19(7):2354–2372

Chapter 15

Enhanced Isolated Boost DC–DC Converter with Reduced Number of Switches



Anjel J and Gerald Christopher Raj I

Abstract This chapter documents a new two-switch, isolated boost DC–DC converter. At the point when the field effect transistor (FET) is turned on, the component produces a resonant pulse, which is then filtered by the output LC like a traditional switching converter. In a quasi-resonant converter (QRC), the width and amplitude of the pulse are fixed and the converter operates under variable frequency by using a resonant pulse, the switching element will naturally go into a zero-current state. To better comprehend the operation of this circuit, we can simplify our model slightly and break up its operation into six modes of intervals. High boost DC–DC converters are required to convert low voltages into a steady DC bus voltage. For galvanic isolation, segregated topologies are generally used. As a result, the voltage-fed full-bridge (VFFB) DC–DC converters are appropriate for applications with high step-up voltage gain. Regarding the performance of the proposed converter, a 343 V DC output was constructed with a 50 V DC input and the higher efficiency obtained is 99%. To check the execution of the proposed converter, a simulation model has to be developed by means of MATLAB Simulink. The developed simulation model needs to be examined under resistive loading condition.

Keywords Voltage-fed full-bridge (VFFB) converter · High-frequency step-up transformer · Voltage double rectifier (VDR)

Abbreviation

CFFB	Current-fed full-bridge
DCM	Discontinuous conduction mode
FC	Fuel cells
PID	Proportional integral derivatives

Anjel J · Gerald Christopher Raj I (✉)

Department of Electrical and Electronics Engineering, PSNA College of Engineering and Technology, Silvarpatti, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_15

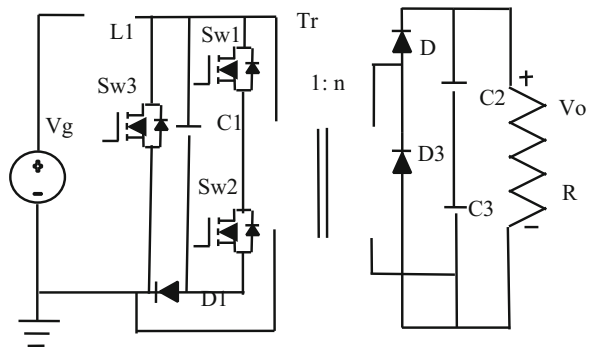
PV	Solar photovoltaic systems
QRC	Quasi-resonant converter
UPS	Uninterruptible power supplies
VDR	Voltage double rectifier
VFFB	Voltage-fed full-bridge
ZCS	Zero current switch
ZVS	Zero voltage switch

15.1 Introduction

15.1.1 A Subsection Sample

High boost DC–DC conversion techniques are required in several applications such as fuel cells (FCs), medical equipments, solar photovoltaic (PV) systems and uninterruptible power supplies (UPS). To improve the voltage boost ability, a boost converter is connected to the secondary side of the voltage-fed full-bridge (VFFB) DC–DC converter. The CFFB converters, in any case, the resonance between the leakage inductor of the transformer and the output capacitance of the primary switches reasons voltage spikes in the devices. A VFFB converter was proposed in order to diminish the input current ripple. The isolated boost converters are used, while the size of the magnetic components and the current stress of the devices are lessened. An input-series output-parallel connection for VFFB converter modules was proposed in order to build the voltage-blocking capability at the input and lessening the current swell at the output. A circuit diagram of the existing three-switch isolated boost DC–DC converter is shown in Fig. 15.1.

Fig. 15.1 Existing three-switch, isolated boost DC–DC converter



15.2 Operating Principle

The half bridge is used for a turned-on/turn-off alternative with a phase shift for conversion of source for each cycle any of the switches is turned on. The transformer is used to step up with a limited gain, to resolve the size issues. The secondary side has a voltage doubler that doubles the input voltage twice. A circuit diagram of the proposed two-switch isolated boost converter is shown in Fig. 15.2.

The LLC resonant converters are proposed, which are well suited for high-voltage and high-frequency applications. To get high-voltage gain, transformer-magnetizing inductance must be small, which increases magnetizing current and higher conduction and core losses. The input boost circuit consists of inductance L_B and Leakage inductance is modelled as L_k . The diodes $D1$ and $D2$ make a full wave rectifier that provides voltage doubler circuit with capacitors C_{o1} and C_{o2} in output. In order to put as in separate like primary transformer turn ratio and secondary transformer turn ratio, ($N1$ and $N2$) and the winding of the transformer (Tr). f_s is the switching frequency and f_r is the resonant frequency.

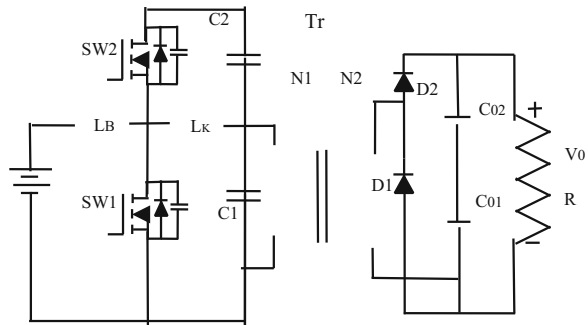
15.2.1 Operating Intervals of Proposed Converter

The operating intervals of quasi-resonant converter are shown in Fig. 15.3.

Stage 1—Interval 1 [t_0 t_1]. At t_0 , the SW2 goes OFF and the transformer current charges and discharges parasitic capacitor of SW2 and SW1, respectively. The voltage across SW1 is zero.

Stage 2—Interval 2 [t_1 t_2]. At t_1 , the current of switch SW1 alters its direction and flows through the switch because SW1 has been turned ON in the past interval. I_{in} circulates through SW1 and V_{in} charges L_B . i_{lk} decreases to zero and the output D_2 is conducting.

Fig. 15.2 Proposed two-switch, isolated boost DC–DC converter



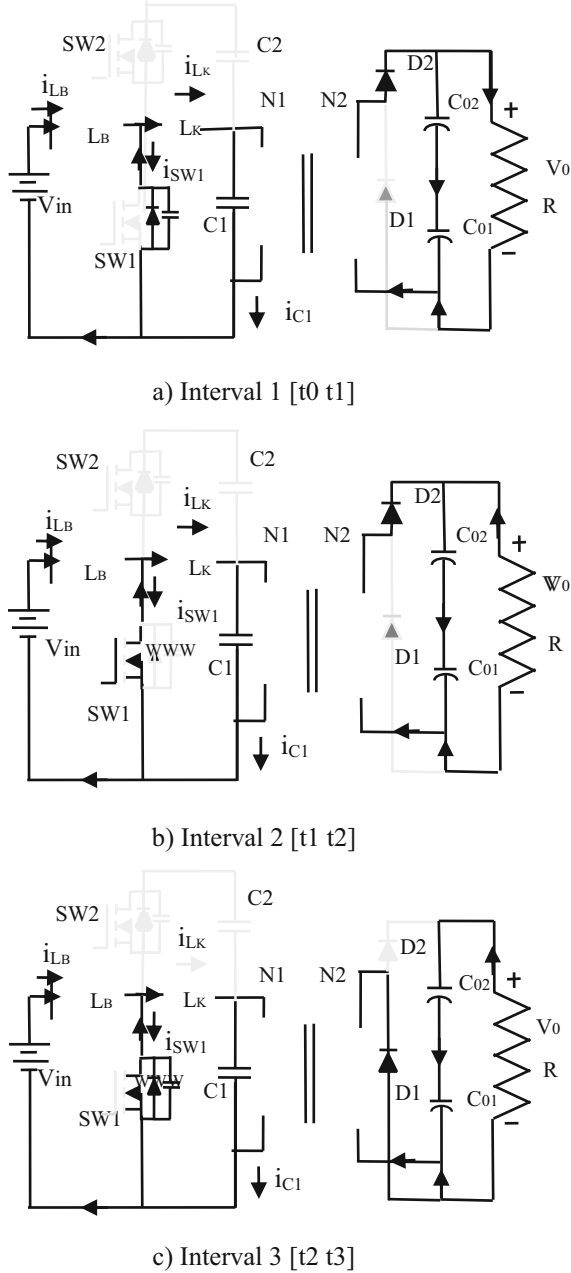
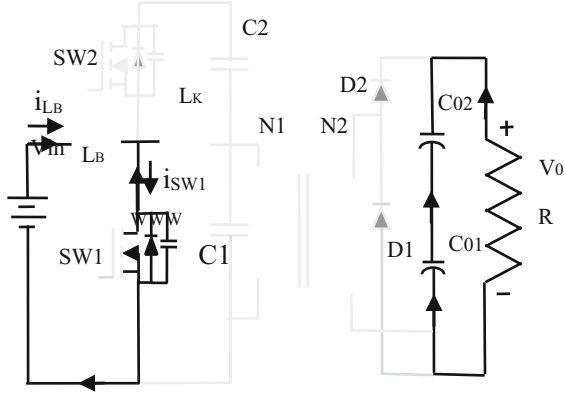
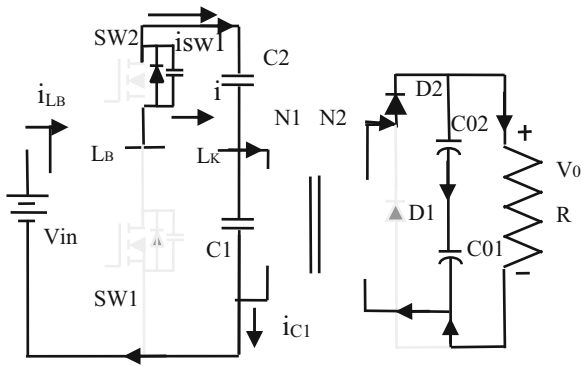


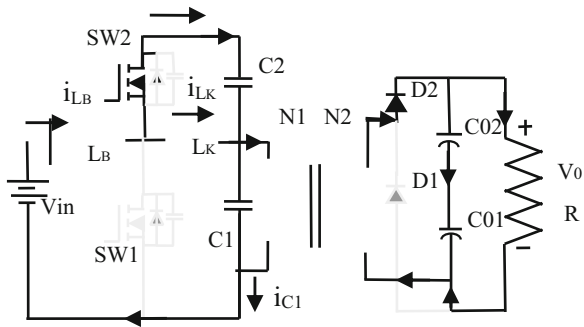
Fig. 15.3 Operating intervals of quasi-resonant converter (a) to (f). (a) Interval 1 [$t_0 t_1$], (b) Interval 2 [$t_1 t_2$], (c) Interval 3 [$t_2 t_3$], (d) Interval 4 [$t_3 t_4$], (e) Interval 5 [$t_4 t_5$], (f) Interval 6 [$t_5 t_6$]



d) Interval 4 [$t_3 t_4$]



e) Interval 5 [$t_4 t_5$]



f) Interval 6 [$t_5 t_6$]

Fig. 15.3 (continued)

Stage 3—Interval 3 [t2 t3]. At t2, the transformer current direction changes and C_{o1} is reflected by means of transformer to the primary side of transformer and C_1 and L_k make resonance. i_{lk} reaches to zero and D_1 is conducting at doubler circuit and supplies C_{o1} .

Stage 4—Interval 4 [t3 t4]. At t3, the current of transformer reaches to zero. The capacitor C_1 and inductor L_k transfer their energy and goes to discontinuous conduction mode (DCM). The power supply charges LB. Then, the output capacitors (C_{o1} and C_{o2}) supply the load.

Stage 5—Interval 5 [t4 t5]. At t4, the SW1 turns OFF. After that the current flows through body diode of switch SW2 and charges C_2 . In output, D_2 is conducting and supplies C_{o2} and the SW2 achieve zero, this interval will wrap up.

Stage 6—Interval 5 [t5 t6]. At t5, the SW2 ranges to zero. The capacitor C_1 has charged due to the fact that t4, the input constant current, flows through the transformer. This interval wraps up by removing pulse gate of SW2. The current of transformer rises linearly from 0 to 1 during the time between t4 and t6.

15.2.2 Voltage Gain

To easily derive the input voltage, the steady state average voltage of the windings of transformer and inductors is zero, so

$$\langle V_{in} \rangle = \langle V_{c1} \rangle; \Delta V_{c1} = \frac{I_{in}}{C_1} (1 - D) T_S \quad (15.1)$$

The Boost gain (M) is given by the equation

$$M = \frac{D}{\left(\frac{2nL_k f_s}{R_L(1-D+d_1)} + (D - d_1) \cdot \frac{1-D}{n} \right)} \quad (15.2)$$

15.3 Simulation Results

In order to validate the performance of the quasi-resonant DC–DC converter, the system is designed with the source modelling in MATLAB/Simulink. For closed loop, V_{ref} is set as 350 V. The PID controller output is constrained to the range of [0.05–0.95]. The closed loop simulated wave form for V_{in}/V_{out} of the quasi resonant is shown in Fig. 15.5. When the input is in 40 V and 60 V, the comparative efficiency curves for the two switches, isolated boost DC–DC converter, are given in Figs 15.6 and 15.7 (Table 15.1).

Fig. 15.5 Input voltage and output voltage of the quasi resonant at closed loop $V_{in} = 50\text{ V}$ and $V_{out} = 343\text{ V}$

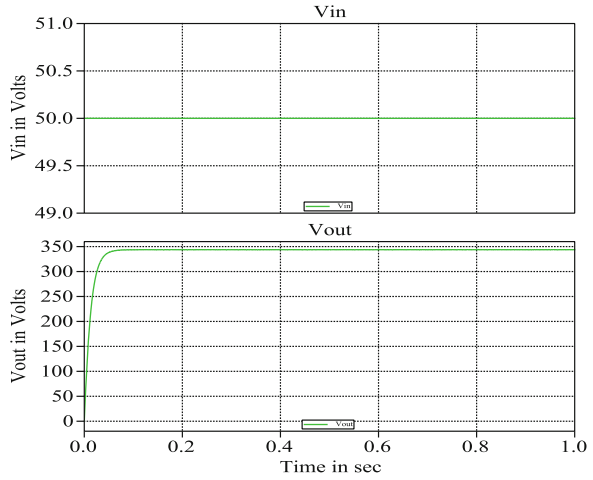


Fig. 15.6 Efficiency of the quasi resonant at closed loop

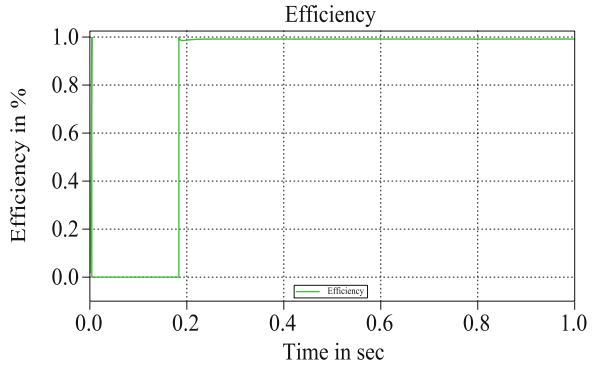


Fig. 15.7 Measured efficiency vs. output power of the quasi resonant at closed loop

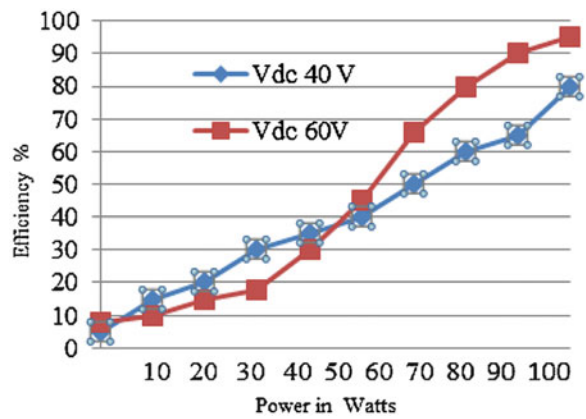


Table 15.1 Simulation parameters for proposed converter

Parameters		Value (Units)
Input voltage (V_{in})		40–60 V
Transformer	Turn ratio	1:N
Switching frequency ($\dot{u}53; \dot{u}60;$)		100 kHz
Resistive load	R_1	120 Ω
	R_2	0.1 Ω
Diode	Forward voltage (V_f)	0.9 V

15.4 Conclusions

An enhanced isolated boost DC–DC converter with two switches was proposed. A 50 V input, 343 V output and rated power model are verified, and the results show that the operation principle of the resonant boost converter with two switches was proposed in this paper and are effective. The efficiency of the proposed converter is 99%. It is higher than that of the existing a three-switch, isolated boost DC–DC converter over wide ranges of output power. The leakage inductance of transformer has been employed to make the resonant circuit with clamp capacitor. The proposed converter achieves ZVS for switches and ZCS for diodes over a wide load range.

References

1. Gautam DS, Musavi F, Eberle W, Dunford WG (2013) A zero-voltage switching full-bridge dc–dc converter with capacitive output filter for plug-in hybrid electric vehicle battery charging. *IEEE Trans Power Electron* 28(12):5728–5735
2. Xie Y, Ghaemi R, Sun J, Freudenberg JS (2009) Implicit model predictive control of a full bridge dc–dc converter. *IEEE Trans Power Electron* 24(12):2704–2713
3. Senthil, Kumar R, Ramesh M, Sundararaju K (2017) Design of Multi-Stack Voltage Equalizer for Partially Shaded PV modules using artificial neural network. *Int J Control Theory Appl* 10:1–19
4. Kumar RS, Kirthika S, Sundararaju K (2017) Analysis of single-stage high-frequency resonant ac/ac converter using artificial neural networks. *Int J Pure Appl Math* 117(8):161–165
5. De Doneker RW, Divan DM, Kheraluwala MH (Jan./Feb 1991) A three-phase soft-switched high-power-density dc/dc converter for high-power applications. *IEEE Trans Ind Appl* 27(1):797–806
6. Ayyanar R, Mohan N (2001) Novel soft-switching DC–DC converter with full ZVS-range and reduced filter requirement—part II: constant-input variable-output applications. *IEEE Trans Power Electron* 16(2):193–200

Chapter 16

Harmonic Intensity Reduction Technique for Three Phase VSI Drive through Double Randomness



P. Arulkumar, K. Jaiganesh, and N. P. Subramaniam

Abstract Deterministic Pulse Width Modulation (PWM) methods are popular in industrial applications due to merits. However, the randomized PWM (RPWM) with its cleaner harmonic spectrum is gaining interest for industrial applications required to meet electromagnetic compatibility standards with almost all the earlier merits retained. The proposed Harmonic Intensity Reduction Double Randomness PWM (HIRDRPWM) technique is a hybrid RPWM, which attains the randomness in two ways. The first one is, in the pre-pulse generation stage, through the chaotic frequency generator, which generates a random frequency carrier (triangular) wave. Second randomness is, in post-pulse generation stage, by varying the position of the pulse. The competence in spreading the harmonic power of sinusoidal PWM (SPWM) and the HIRDRPWM is compared using simulation. The distribution of harmonic power in the output voltage of VSI with induction motor load is studied using the MATLAB software. The discussion includes Total Harmonic Distortion (THD) in output line voltage and the Harmonic Spread Factor (HSF).

Keywords Harmonic Intensity Reduction Double Randomness Pulse Width Modulation (HIRDRPWM) · Harmonic Spread Factor (HSF)

Abbreviation

ASD	Adjustable-speed drive
CCSPWM	Chaotic carrier sinusoidal PWM
EMI	Electromagnetic interference

P. Arulkumar (✉)

Balaji Institute of Technology and Science, Narsampet, Warangal, Telangana, India

K. Jaiganesh

Vardhaman College of Engineering, Hyderabad, Telangana, India

N. P. Subramaniam

Pondicherry Engineering College, Puducherry, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_16

HIRDRPWM	Harmonic intensity reduction double randomness PWM
HSF	Harmonic spread factor
PRBS	Pseudorandom binary sequence
PSD	Power spectral density
PWM	Pulse width modulation
RPWM	Random pulse width modulation
SPWM	Sinusoidal PWM
THD	Total harmonic distortion
VSI	Voltage source inverter

16.1 Introduction

A voltage source inverter (VSI) is the versatile AC drive, which offers a variable frequency and variable voltage, and established as dominant adjustable-speed drive in almost all industrial applications. Sinusoidal pulse width modulation (SPWM) technique has become the most popular and important pulse width modulation (PWM) technique for VSI-based drive systems. PWM-VSIs are the dominant power conversion system in industry today [1–4]. Ahead of the SPWM, a large number of PWM switching pattern generators have been developed over the last four decades to meet the respective requirements for “clean” output waveforms of sinusoidal nature. The choice among the PWM methods depends on the application [5–10].

The deterministic PWM in the inverter drive systems, however, results in the concentration of the output power harmonics at discrete frequencies at the PWM switching frequency and multiples of it. This results in objectionable acoustic noise, electromagnetic interference (EMI), vibration and harmonic heating. If the randomness is introduced in the pulse generation, the harmonics content will spread over wide range and the specific harmonic parts can be significantly reduced. This is the principle of random PWM techniques, which have received much attention very recently [11, 12].

A new hybrid random pulse width modulation (PWM) scheme has been proposed in order to disperse the acoustic switching noise spectra of an induction motor drive [13]. The proposed random PWM pulses are produced through the logical comparison of a pseudorandom binary sequence (PRBS) bits with the PWM pulses corresponding to two random triangular carriers. A constant frequency approach, which has gained popularity by introducing the randomness in the position of switching pulses within switching cycles, has been studied [14]. Yash Shrivastava et al. have suggested a statistical approach to the analysis of random pulse width modulation methods that generate PWM signals by comparing a reference modulating function with random numbers or signals [15–17]. The harmonic spreading effects of existing RPWM methods are not appreciable due to their simple random sequence/logic. A digitally controlled double randomness pulse width modulated voltage source inverter, possessing the accomplishment of distributing the harmonic

power, is proposed for induction motor drive application. The proposed Harmonic Intensity Reduction Double Randomness PWM (HIRDRPWM) technique is a hybrid RPWM, which attains the randomness in two ways. The first one is, in the pre-pulse generation stage, through the chaotic frequency generator, which generates a random frequency carrier (triangular) wave. Second randomness is, in post-pulse generation stage, by varying the position of the pulse.

The chaotic sequence can always be employed for incorporating high degree of randomness. The harmonic spreading effects of sinusoidal PWM (SPWM) and the HIRDRPWM are compared using MATLAB simulation. The discussion includes total harmonic distortion (THD) in output line voltage and the harmonic spread factor (HSF).

16.2 Study of Harmonic Distribution in SPWM VSI Drive

Figure 16.1 represents the general structure of PWM-VSI-fed induction motor (IM) drive, which consists of front-end rectifier, DC link filtering components, VSI and the motor. In the basic SPWM, a fixed frequency triangular carrier is compared directly with a sinusoidal reference, and the resulting signals are fed to switches. The conventional deterministic PWM in the VSI drive systems, however, results in the concentration of the output power harmonics at discrete frequencies related to the fixed switching frequency [16]. These harmonic powers may cause the undesired electromagnetic noise and psychoacoustic noise for human beings.

The HSF quantifies the spread spectra effect of the random PWM scheme. For this purpose, the concept of statistical deviation can be employed and the HSF [17] is defined as follows:

$$HSF = \sqrt{\frac{1}{N} \sum_{K=0}^N (H_j - H_0)^2} \tag{16.1}$$

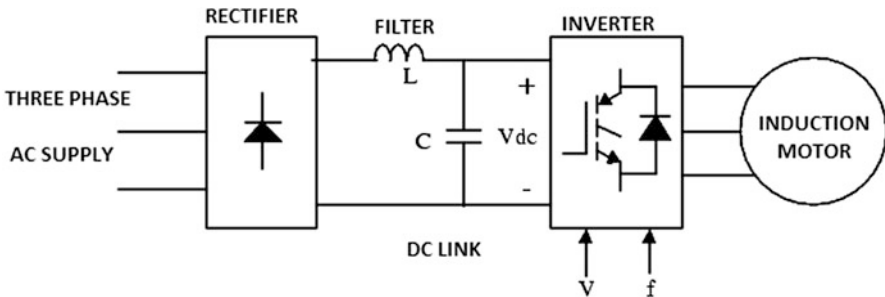


Fig. 16.1 IGBT-based VSI-fed IM drive

$$H_0 = \sum_{j>1}^N (H_j) \tag{16.2}$$

Where ‘ H_j ’ is amplitude of j th harmonics and ‘ H_o ’ is average value of all ‘ N ’ harmonics. Understanding how the strength of a signal is distributed in the frequency domain, power spectral density (PSD) graphs are used.

The harmonic spreading effects of conventional SPWM are evaluated in this section. The simulation study is performed in MATLAB/Simulink software. A three-phase VSI inverter with induction motor load is considered. The input DC voltage (V_{DC}) is 415 V and the output frequency is taken as 50 Hz. The carrier frequency (f_c) is taken as 3 kHz. The load is a three-phase squirrel cage induction motor load (0.75 kW and 2.5 A) and ODE Solver ode23tb is used [18, 19]. The line voltage waveform resulted from SPWM is illustrated in Fig. 16.2 for $M_a = 0.8$ and its corresponding harmonic spectrum is shown in Fig. 16.3.

The THD, HSF and fundamental component (V_o) of the output voltage are listed for the complete working range in Table 16.1. The variation HSF with respect to modulation index is shown in Fig. 16.4. The linear relation between V_o and M_a , and indirect proportionality of THD with M_a are studied. The variation of HSF with M_a is an interesting result and worth to note. Both THD and HSF are decreased at higher M_a values. The HSF at $M_a = 0.8$ is 5.566.

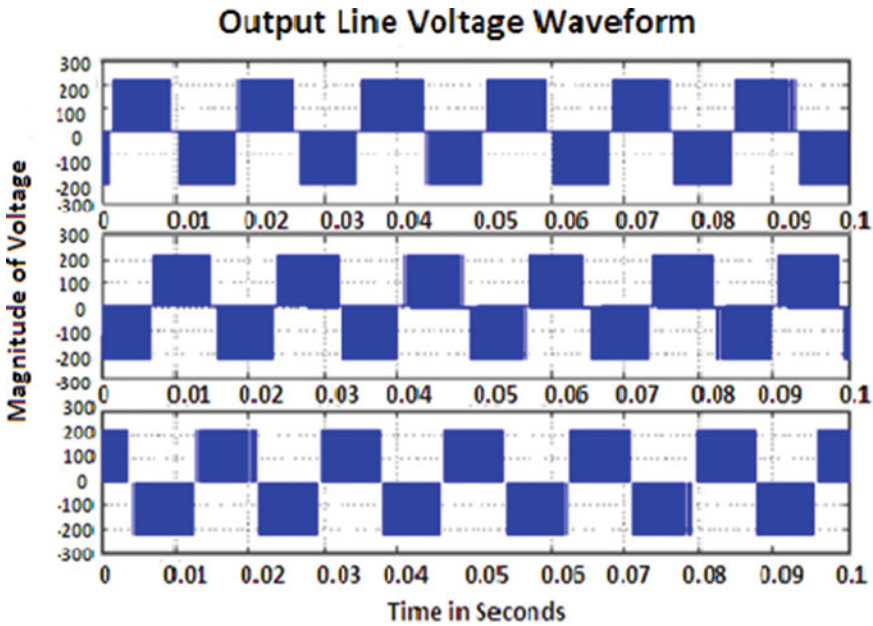


Fig. 16.2 Line voltage waveform – SPWM

Fig. 16.3 Harmonic spectrum of line-line voltage – SPWM

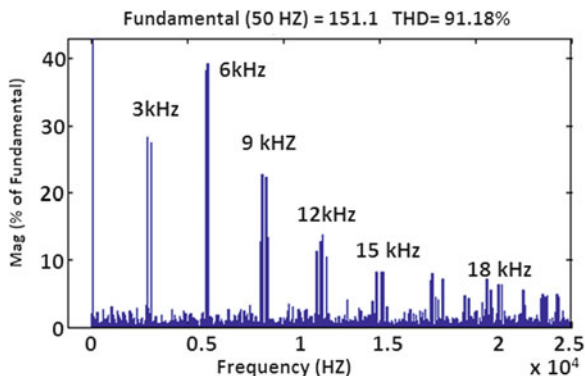


Table 16.1 Performance of SPWM

M_a	V_o (V)	THD %	HSF
0.2	70	257.97	8.312
0.4	135	164.31	6.142
0.6	210	121.10	5.880
0.8	287	90.60	5.566
1.0	357	81.01	5.20
1.2	393	68.30	5.00

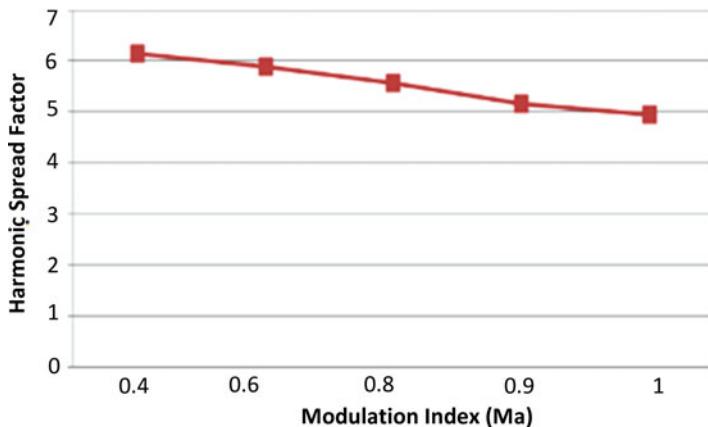


Fig. 16.4 Harmonic spread factor versus modulation index – SPWM

16.3 Proposed HIRDRPWM

RPWM switching is an effective way to let the harmonic spectrum of a power converter be uniformly distributed. Conceptually, RPWM can be made by randomly varying the key properties of the PWM mechanism, such as frequency or triangular carrier slope, switching pulse position or other specific attributes.

The basic principle of the proposed HIRDRPWM is described in Fig. 16.5. In developed scheme, the randomness is added at both the pre-pulse generation and the

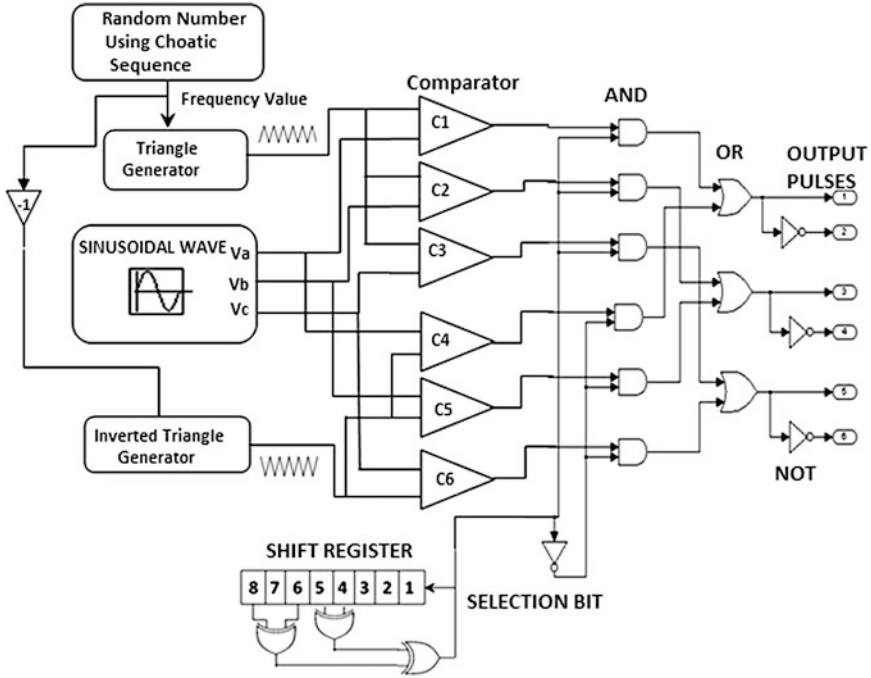


Fig. 16.5 HIRDRPWM scheme

post-pulse generation stages. A chaotic algorithm generates the triangular signal with random frequency by a random number, which decides the carrier frequency of next cycle. The basic principle of CCSPWM is to use a chaotic signal to vary the switching (or carrier) frequency. The following chaotic function is involved in generating the variable frequency carrier signal.

$$f_n = f_{low} + (f_{high} - f_{low} + 1) \frac{x_n}{0.5(5^c - 1)} \tag{16.3}$$

where f_n is the n -th switching frequency of chaotic PWM and chaotic sequences x_n may be generated simply by iteration. Thus, the switching frequency may be varied from f_{low} to f_{high} . The constant c is assumed as 6.

After having decided on the carrier cycle, the pulses are generated for all the three phases by comparing the conventional sinusoidal reference with the generated carrier (+triangular). Similar comparison is also done for inverted version of the triangular wave (−triangular). The second randomness is incorporated with these six pulses in process of selecting three phases. The selection of among these two groups of pulses is done using a select signal, pseudorandom binary sequence (PRBS) bits. If the PRBS bit is 1, pulses of group 1 are selected else group 2. Once the group is selected, then pulses of lower arm devices are generated by inversion.

16.4 Simulation Study

The simulated waveforms of three-phase motor line voltages and currents are represented in Figs. 16.6 and 16.7, respectively. The harmonic spectrum and the PSD for $M_a = 0.8$ and 1.2 are diagrammed as Figs. 16.8 and 16.9, respectively.

From Table 16.2, it is understood that THD and HSF of the proposed HIRDRPWM are reduced for the entire range of M_a . At the modulation depth of 0.2, the reduction HSF is about 51% in the HIRDRPWM. The percentage reduction in HSF is more at lower modulation depths in linear modulation and also at higher depths in over-modulation region.

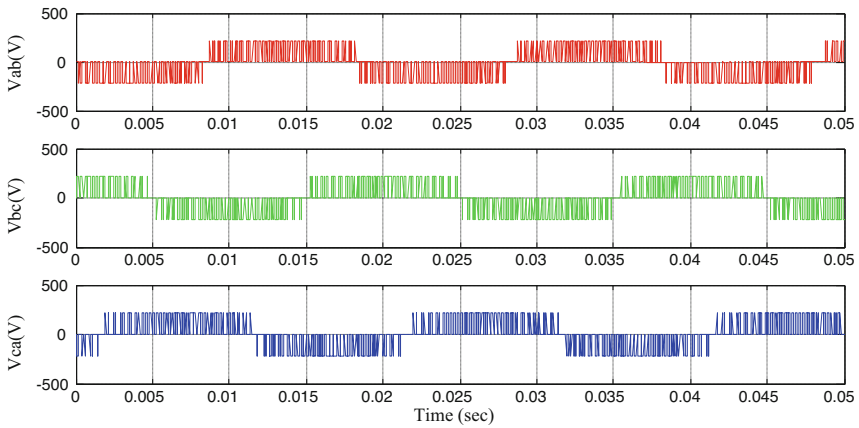


Fig. 16.6 Simulated line-line voltage waveform for $M_a = 0.8$

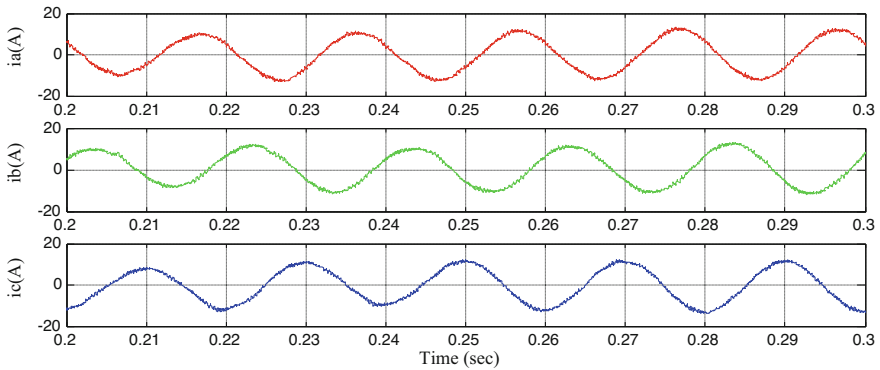


Fig. 16.7 Simulated line current waveform for $M_a = 0.8$

Fig. 16.8 Line voltage harmonic spectrum

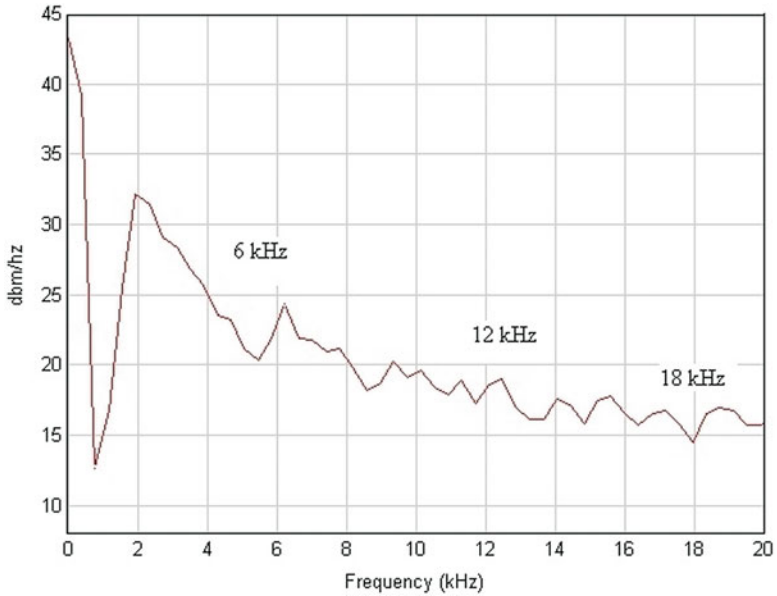
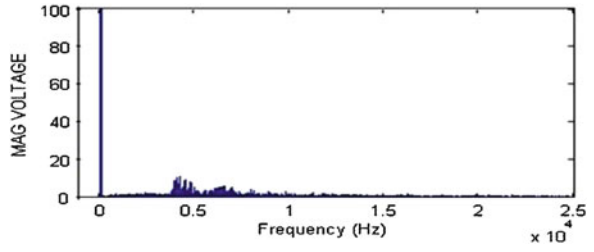


Fig. 16.9 Power spectral density

Table 16.2 Performance comparison of SPWM and HIRDRPWM

M_a	THD (%)		HSF		Reduction HSF (%)
	SPWM	HIRDRPWM	SPWM	HIRDRPWM	
0.2	257.97	252.52	8.31	4.04	51.34
0.4	164.31	160.54	6.14	3.41	44.46
0.6	121.10	118.79	5.88	3.24	44.90
0.8	90.60	86.21	5.57	2.81	49.55
1.0	81.01	63.21	5.20	2.22	55.15
1.2	68.30	51.11	5.00	2.05	56.66

16.5 Conclusion

Even the prime advantage of SPWM technique is having an assortment of performances such as high output quality, less THD, low rating filtering requirements and linear control on fundamental component etc. The harmonic components of the output voltage are concerted around switching frequency and its integer multiples. These distinct dominant harmonics result mainly in torque ripples and acoustic noise in drives. These problems are mitigated by spreading the harmonic power in the output voltage through non-deterministic pulse width modulations. The proposed harmonic intensity reduction double randomness PWM (HIRDRPWM) technique is a hybrid RPWM, which attains the randomness in two ways. The distribution of harmonic power in the output voltage of VSI with induction motor load is studied using MATLAB software. The discussion includes total harmonic distortion (THD) in output line voltage and the harmonic spread factor (HSF). The simulation study reveals that the proposed scheme helps in reducing the HSF about 50%. This offers the reduction in acoustic noise and vibration in ASDs.

References

1. Chiasson J, Tolbert LM, McKenzie K, Du Z (2004) A complete solution to the harmonic elimination problem. *IEEE Trans Power Electron* 19(2):491–499
2. Mohan N, Undeland TM, Robbins WP (1995) *Power electronics: converters, applications, and design*, 2nd edn. Wiley, New York
3. Holtz J (1992) Pulse width modulation—a survey. *IEEE Trans Ind Electron* 39(5):410–420
4. Van Der Broeck HW, Skudelny HC, Stanke GV (1988) Analysis and realization of a pulse-width-modulator based on voltage space vectors. *IEEE Trans Ind Appl* 24:142–150
5. Iqbal A, Ahmed SM, Khan MA, Abu-Rub H (2010) Generalised simulation and experimental implementation of space vector PWM technique of a three-phase voltage source inverter. *Int J Eng Sci Technol* 2(1):1–12
6. Solomon OR, Famouri P (2006) A novel approach for evaluating performance of discontinuous pulse width modulation schemes for three-phase voltage source inverter. In: The proceedings of IEEE conference (ISIE 2006), July 2006, Montreal, Quebec, Canada, pp 9–12
7. Houlds Worth JA, Grant DA (1984) The use harmonic distortion to increase the output voltage of three-phase PWM inverter. *IEEE Trans Ind Appl* 1 IA-20:1224–1228
8. Ziogas PD (1981) The delta modulation technique in static PWM inverters. *IEEE Trans Ind Appl* 1A-17:199–203
9. Lynn Kirlin R, Kwok S, Legowski S, Trzynadlowski AM (1994) Power spectra of a PWM inverter with randomized pulse position. *IEEE Trans Power Electron* 9(5):463–472
10. Das S, Narayanan G (2012) Novel switching sequences for a space-vector-modulated three-level inverter. *IEEE Trans Ind Electron* 59(3):1477–1487
11. Khan H, Miliiani E-H, Drissi KEK (2012) Discontinuous random space vector modulation for electric drives: a digital approach. *IEEE Trans Power Electron* 27(12):4944–4951
12. Mathe L, Lungeanu F, Sera D, Rasmussen PO, Pedersen JK (2012) Spread Spectrum modulation by using asymmetric-carrier random PWM. *IEEE Trans Ind Electron* 59(10):3710–3718
13. Kim K-S, Jung Y-G, Lim Y-C (2009) A new hybrid random PWM scheme. *IEEE Trans Power Electron* 24(1):192–200

14. Kirlin RL, Kwok S, Legowski S, Trzynadlowski AM (1994) Power spectra of a PWM inverter with randomized pulse position. *IEEE Trans Power Electron* 9(5):463–472
15. Shrivastava Y, Sathiakumar S, (Ron) Hui SY (1998) Improved spectral performance of random PWM schemes with weighted switching decision. *IEEE Trans Power Electron* 13 (6):1038–1045
16. Tanaka T, Ninomiya T, Harada K (1989) Random-switching control in DC-to-DC converters. *IEEE PESC Rec* 1:500–507
17. Legowski S, Bei J, Trzynadlowski AM (1992) Analysis and implementation of a grey-noise PWM technique based on voltage space vectors. In: *Proceedings of IEEE international applied power electronics conference (APEC-1992)*, pp 586–593
18. Arulkumar P, Subramaniam NP (2015) Chaotic triangular carrier based non-deterministic SPWM strategy for voltage source inverter drives. *Indian J Sci Technol* 8(9):842–848
19. Arulkumar P, Ravichandran M, Subramaniam NP (2015) Analysis of FPGA based non-deterministic PWM in induction motor drives. *Int J Appl Eng Res* 10(4):10441–10452

Chapter 17

PV-Based Multilevel Inverter-Fed Three-Phase Induction Motor with Improved Time and Speed of Response



Chandrasekaran S and Durairaj S

Abstract In the present power scenario, the power quality is most significant in the field of grid-connected and load-connected inverter. A number of techniques were proposed in the field of renewable energy for improving the efficiency and quality of the power. In this article, the time response of the controller was analysed and verified for load-connected multilevel inverter. There are different types of controllers in use, such as Proportional (P), Proportional Integral (PI), Proportional Integral Derivative (PID), Integer Order PID (IOPID), etc. Out of those controllers classic PI controller is most efficient in the speed of response, that's why the classic PI controller is compared with proposed Fractional Order Proportional Integral Derivative controller (FOPID). Based on the results, FOPID has more speed of response. The outputs of the multilevel inverter are to be improved by minimizing the rise time, settling time and steady-state error of the inverter. The Simulink model was built to control the motor speed and it will going to be applied for grid in the future.

Keywords PID · Fractional-order PID · MLI · Harmonic · Renewable energy · Total Harmonic Distortion (THD)

Abbreviation

DC	Direct Current
FOPID	Fractional-order Proportional Integral Derivatives
HPWM	Hybrid Pulse Width Modulation
IOPID	Integer-order Proportional Integral Derivatives
MLI	Multilevel Inverter

Chandrasekaran S (✉)

Arasu Engineering College, Chennai Main Road, Kumbakonam, Thanjavur, Tamil Nadu, India

Durairaj S (✉)

Dhanalakshmi Srinivasan Engineering College, Thuraiyur Road, Perambalur, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_17

MOSFET	Metal-oxide Semiconductor Field Effect Transistor
PI	Proportional Integral
PID	Proportional Integral Derivatives
PV	Photo-voltaic
RPM	Revolution Per Minute
RPWM	Random Pulse Width Modulation
SPWM	Single Pulse Width Modulation
THD	Total Harmonic Distortion
VCO	Voltage-controlled Oscillator

17.1 Introduction

The photovoltaic panel has played an important role in the field of renewable energy sector. There are a number of disturbances which oppose the quality of power. This article is focused on the speed response of the system. The classic PI controller is mainly used for industrial application. It is used for under damped control systems [1]. The speed of response of the proportional integral (PI) controller was very poor; in turn the stability of the system is also negative. This is mainly used for the industrial applications. This controller is mainly suitable for controlling the nonlinear boost converter [2].

The cascaded H-bridge MLI can be able to produce more accurate output when compared with classical inverters [3, 4]. The three-phase induction motor acts as a load on the system. In the output side of the system, the speed of the motor is measured and it is given as an input to the controller.

In this article, buck–boost converter operations are discussed in Sect. 17.2, principles and operation of cascaded multilevel inverter are illustrated in Sect. 17.3, Section 17.4 represents the response of PI controller; construction and applications of proposed system with fractional-order proportional integral derivative (FOPID) controller and time response are discussed in Sect. 17.5; harmonic reductions are analysed in Sect. 17.6, the simulation results are verified in Sect. 17.7 and comparisons of output parameters are carried out in Sect. 17.8 (Fig. 17.1).

17.1.1 Specification of the Systems

The specification for the existing systems and proposed systems are mentioned in Table 17.1.

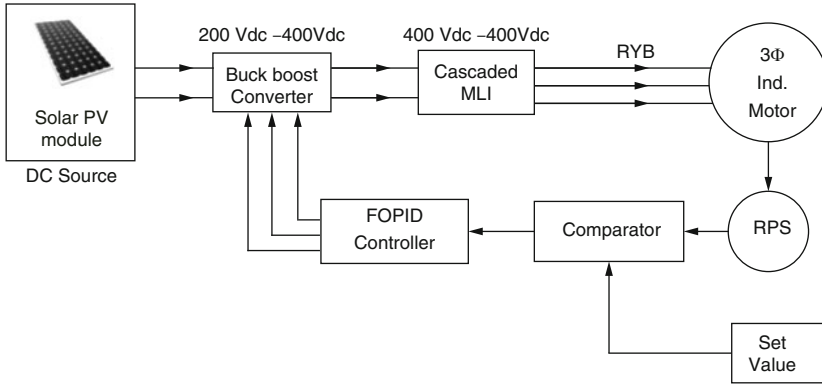


Fig. 17.1 Proposed systems

Table 17.1 System specifications

Buck-boost converter	Inverter	Three-phase induction motor	PI	FOPID
$R = 0.001$ ohms	$R =$ ohms	Voltage = 460 V	$K_p = 0.18$	Elliptic filter order = 5
$L = 2e^{-3}$ H	$R = 0.001$ ohms	$N = 1750$ rpm	$K_i = 8$	$K_p = 0.9, K_i = 9, K_d = 0.0009$
$C = 50e^{-6}$ F	$L = 1e^{-6}$	$H_p = 5$	Time = $50e^{-6}$	VCO, amp = 50 V

17.1.2 Sizing of the Photovoltaic Panel

The photovoltaic module sizing is designed and calculated based on the following parameters: electrical load, inverter specification, battery size and specification, PV panel array specification and controller specification [5]. For the panel sizing, the total watt-hour per day is needed. After calculating the watt-hour per day, it is multiplied by 1.3 (loss in the system).

17.2 Buck-Boost Converter

The ideal switch acts as a dc source. The switch consists of 200 V_{dc} source and 8 V_{dc} is disturbance voltage. The metal-oxide semiconductor field-effect transistor (MOSFET) has 360° operations. It has more advantages. The circuit parameters such as resistance $R = 0.001$ ohms, capacitance $C = 50e^{-6}$ and inductance $L = 2e^{-3}$ are used. A number of Pulse Width Modulation (PWM) techniques are employed in the converter and inverter circuit. There are Single Pulse Width Modulation (SPWM), Random Pulse Width Modulation (RPWM), Hybrid Pulse Width Modulation (HPWM) inverter [6–8] and buck converter, boost converter, buck-boost converter, etc.,

17.3 Cascaded Multilevel Inverter

The inverter circuit consists of eight MOSFET switches M1, M2, M3, M4, M5, M6, M7, M8 for each phase. This circuit was designed for three-phase supply. All the switches work on the bases of operation of the pulse generator. The MOSFET has different parameters such as diode resistance, diode inductance, snubber resistance, snubber capacitance, etc.

The phase to neutral voltage can be obtained from

$$\begin{aligned} V_{an} &= V_m \cos \omega t, \\ V_{bn} &= V_m \cos \left(\omega t - \frac{2\pi}{3} \right), \\ V_{cn} &= V_m \cos \left(\omega t + \frac{2\pi}{3} \right), \\ V_{ab} &= V_{an} - V_{bn} = \sqrt{3} V_m \cos \left(\omega t - \frac{\pi}{6} \right), \\ V_{bc} &= V_{bn} - V_{cn} = \sqrt{3} V_m \cos \left(\omega t - \frac{\pi}{2} \right). \end{aligned}$$

Based on the necessity and quality of power, the levels of inverter may be improved. The levels of inverter are directly proportional to the quality of the inverter (pure sine wave). In this system, seven levels of inverter circuit are used with PI and FOPID controller. The seven steps are produced at the output of the inverter. This inverter is used to reduce the stress of the devices.

17.3.1 Feedback Circuit

The speed reference is taken as a feedback signal from the three-phase induction motor. The feedback is varied on the basis of the gain of the output. The gain for speed reference is; gain = $4/\text{Pi}$ and for electromagnetic torque is; gain = $825/\text{Pi}$ (Fig. 17.2).

17.4 Response of PI Controller

There are different controllers used for improving the quality of the power. Out of all controllers, PI and FOPID controller outputs are used for comparison. The controller is operated on the basis of gain value. The speedy response and overall stability of the system is very poor in the PI controller.

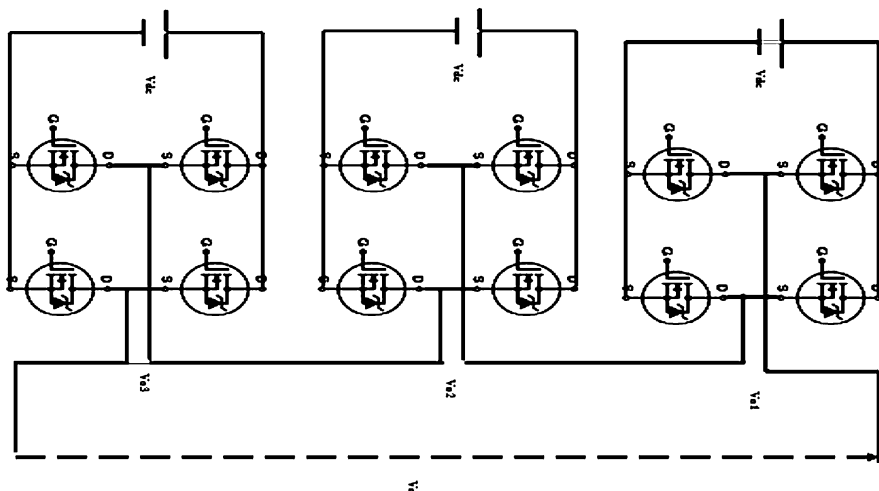


Fig. 17.2 Cascaded seven-level multi-level inverter

17.4.1 PI Controller

The PI controller does not reduce the rise time and the oscillation of the system. The PI controller is functioning on the basis of input parameters such K_p , K_i , output limit and sampling time. The controller parameters are proportional gain $K_p = 0.18$, integral gain $K_i = 8$, output limit (upper limit = 1000 rpm and lower limit = -1000 Rpm) and sampling time, $t = 50e^{-6}$ (Fig. 17.3).

17.5 Proposed System with FOPID Controller and Time Response

In the proposed system, the fractional-order proportional integral derivative controller is used in this system. The FOPID controller operates on the principles of fractional calculus. This controller is used in both power system and power electronics field. The upper output limit is 280 rpm and lower output limit is -280 rpm and sample time is $50e^{-6}$.

17.5.1 FOPID Controller

The FOPID controller is designed using the following parameters and values. It consists of elliptic filter, gain, voltage control oscillators and Proportional Integral Derivative (PID) blocks. The elliptic filter is analogue filter. The fractional values are

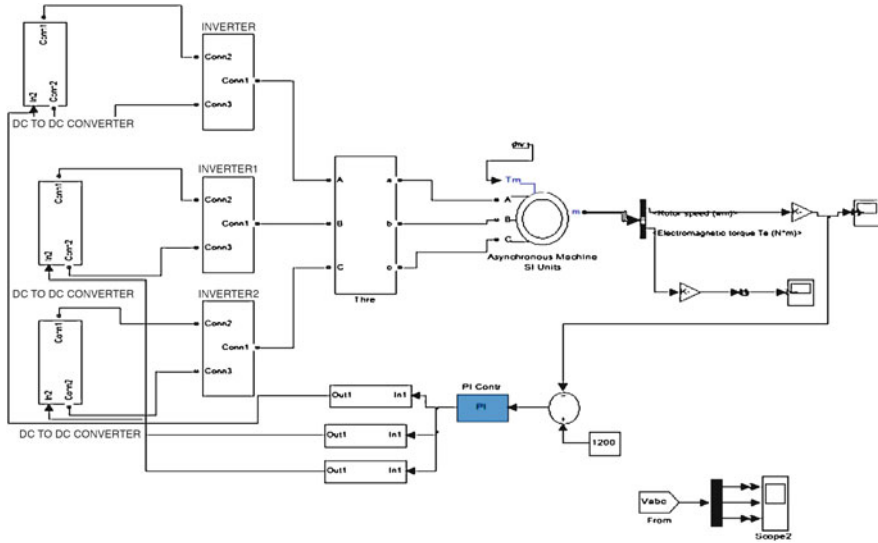
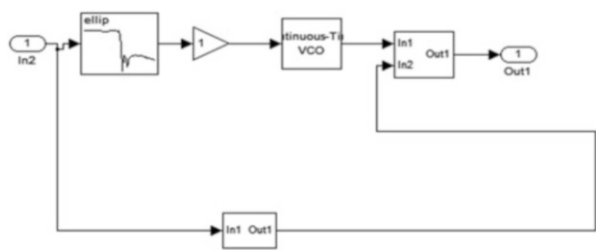


Fig. 17.3 PI controller

Fig. 17.4 FOPID controller



applied to the analogue filter such as filter order = 5, pass band edge frequency = 50 Hz, pass band ripple = 2 and stop band attenuation is 40. In this controller, the time response is more.

The Voltage-controlled Oscillator (VCO) will decide the amplitude of a signal. The PID blocks have the gain values. The output of the controller; controls the input given of the dc–dc converter. The rise time, peak time, settling time and steady state error are reduced and the efficiency is improved (Figs. 17.4 and 17.5).

17.6 Harmonic Reduction

The harmonics are one of the major issues in the power system. In this chapter, the harmonics are analysed for the MLI-fed three-phase induction motor. The magnitude and the THD values of PI controller are; fundamental magnitude (50 Hz) = 19.29

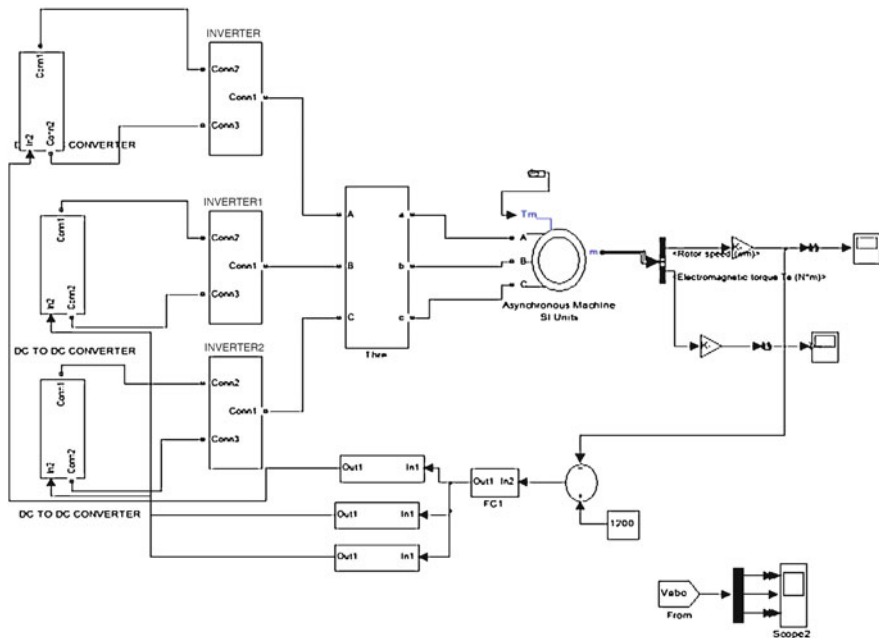


Fig. 17.5 Closed loop with FOPID controller

Table 17.2 Comparison of current harmonics

S. No	Controller	Harmonics values
1.	PI	19.29
2.	FOPID	11.25

and THD = 17.04%. The magnitude and THD values of FOPID controller are; fundamental magnitude (50 Hz) = 11.25 and THD = 16.72%. So THD is also improved.

$$THD = \sqrt{\frac{\sum_{n=3,5,7,\dots} V_n^2}{V_1^2}}$$

17.6.1 Comparison of Harmonics Parameter

In this chapter, current harmonics of existing and proposed systems were compared and the values of the current harmonics were reduced (Table 17.2).

17.7 Simulation Results

17.7.1 Closed Loop with PI Controller

The simulation circuits are designed and the input is given as per the requirement of output voltage of the inverter. In this chapter, the circuit is analysed for closed loop operation. The input voltage is 200 V_{dc} supply and the disturbance voltage is 8 V. The disturbance is used to verify the steady-state stability of the system (Figs. 17.6 and 17.7).

17.7.2 Closed Loop with FOPID Controller

The time response of the FOPID controller is compared to PI controller. The input voltage and disturbance voltage are same as PI controller. The buck–boost converter boosts the voltage in the 1:2 ratios. The input voltage is same as that of PI controller (Fig. 17.8).

The output voltage of inverter with FOPID controller has more efficiency when compared with the PI controller (Figs. 17.9 and 17.10).

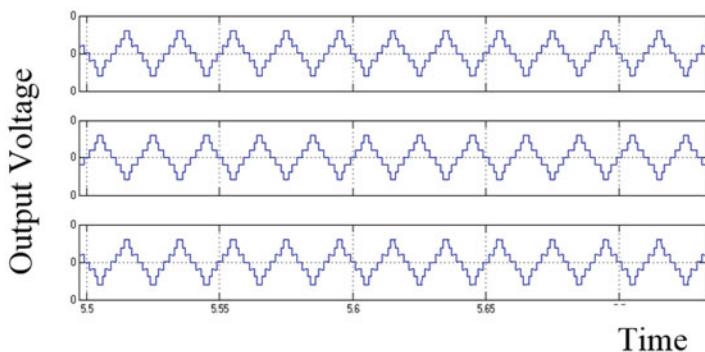
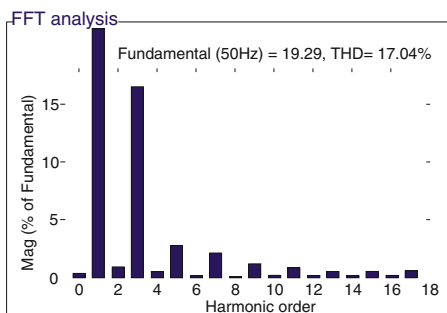


Fig. 17.6 Output voltage of three-phase inverter

Fig. 17.7 Output current THD PI controller (frequency = 19.29, THD = 17.04%)



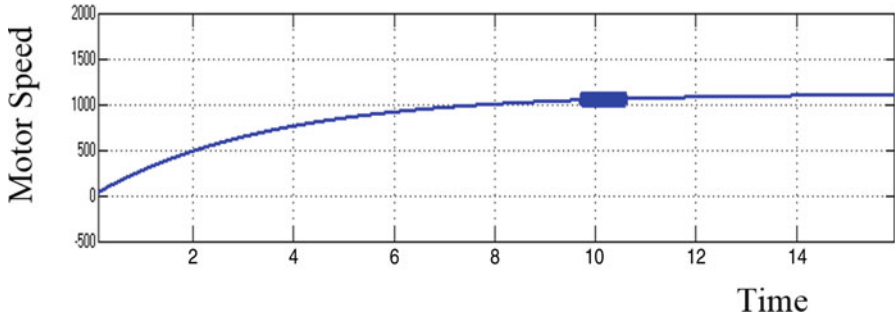


Fig. 17.8 Motor speed

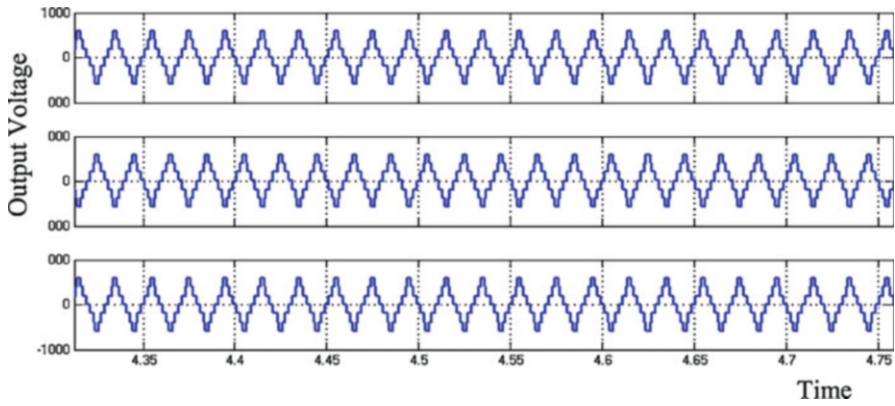


Fig. 17.9 Output voltage of inverter

Fig. 17.10 Output current
THD FOPID controller
(frequency = 11.25,
THD = 16.72%)

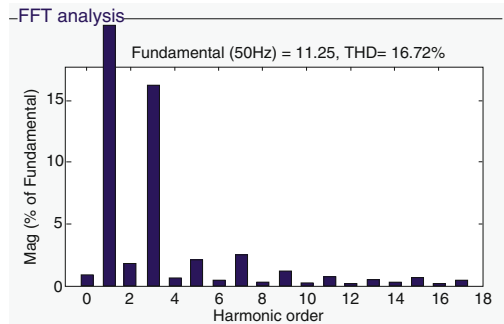


Table 17.3 Comparison of time domain parameters

Controller	Rise time (s)	Peak time (s)	Settling time (s)	Steady state error (RPM)	ISE	Frequency (50 Hz)	THD
PI	9.9	10.3	11.3	8.2	106	19.29	17.04
FOPID	9.7	9.9	10.6	5.8	58	11.25	16.72

17.8 Comparison of Output Parameters

The speed response of both PI controller-based inverters is compared with the FOPID controller. The time and speed response of FOPID controller is improved when compared with PI controller. The overall efficiency of the system also improved (Table 17.3).

17.9 Conclusion

This article represents the analysis of speed of response for the PV-based, seven-level inverter circuit with three-phase induction motor. The outputs of the motor are measured and verified with two different controllers such as PI and FOPID controller. Based on the analysis and result, the FOPID controller produces the more efficient, reduced error and qualified output waveform. The output of the multilevel inverter is verified and quality of the sine wave is improved and compared with original sine waveform. The FOPID controller-based MLI produces the more quality output with reduced distortion when compared with the PI controller. This research may be extended with hysteresis controller and fuzzy logic controller for load-connected PV inverter as the future work.

References

1. Smriti Rao K, Mishra R (2014) Comparative study of P, PI and PID controller for speed control of VSI-fed induction motor. *Int J Eng Dev Res* 2(2):2321–9939
2. Tepljakov A, Alagoz BB, Yeroglu C, Gonzalez E, Nia SHH, Petlenkov E (2018) FOPID controllers and their industrial applications: a survey of recent results. 3rd IFAC conference on advances in proportional-integral-derivative control, Ghent, Belgium, May 9–11, 2018
3. Venkateswara rao R, Ramash Kumar K, Narasimha Raju VSN (2017) Design of PI controller for seven level symmetrical MLI with minimal quantity of switches plus Snubber circuit. *Int J Eng Trends Technol* 47(8):445–452
4. Ahamed M, Sheir A, Orabi M (2017) Real time solution and implementation of selective harmonic elimination of seven-level multi level inverter. *IEEE J Emerg Sel Top Power Electron* 5(4):1700–1709
5. Salam MA, Aziz A, Alwaeli AHA, Kazem HA (2013) Optimal sizing of photovoltaic systems using HOMER for Sohar, Oman. *Int J Renew Energy Res* 3(2)

6. George DS, Baiju MR. Random PWM scheme for 3-level inverter using offset time randomization. IECON 2011 – 37th annual conference of the IEEE industrial electronics society
7. Kulkarni A, John V (2013) Mitigation of lower order harmonics in a grid connected single phase PV inverter. IEEE Trans Power Electron 28(11):5024–5037
8. Ramanathan T, Ramesh R, Subramaniam CK, Ganesan K. Pseudorandom carrier based subharmonic PWM for cascaded multilevel inverters. IEEE international symposium on circuits and systems (ISCAS)

Chapter 18

Adaptive Disturbance Observers for Building Thermal Model



Mallikarjun Soudari, Seshadhri Srinivasan, and B. Subathra

Abstract Space cooling in buildings is influenced by thermal dynamics, which in turn is affected by ambient conditions, solar radiation, occupancy, stray heating and various other disturbances that are time-varying and nonlinear. This investigation presents an adaptive disturbance observer for estimating the thermal states of the building depending on the disturbance influences. In our approach, the building is modelled as an electrical equivalent circuit and the disturbance influences are modelled as exogenous inputs. Then an adaptive observer is designed for estimating the disturbances and providing accurate state estimates. Further, we also provide the conditions in which the adaptive observer provides an estimate of the disturbance. The proposed approach is illustrated on a test building with an air conditioner controlled using a thermostat. Our studies showed that the proposed observer provided accurate estimation of temperature depending on the disturbances.

Keywords Adaptive disturbance estimator · Thermal model · Disturbance observer

Abbreviations

ADE	adaptive disturbance estimator
BCTVB	building controls virtual test bed
BDA	building design advisor
EKF	extended Kalman filter

M. Soudari · B. Subathra
Department of Instrumentation and Control Engineering, Kalasalingam Academy of Research and Education, Krishnan Koil, India

S. Srinivasan (✉)
Berkeley Education Alliance for Research in Singapore, Singapore, Singapore

IDA	indoor climate and energy
MPC	model predictive control
SVR	support vector machines
TRNSYS	transient system simulation tool

18.1 Introduction

Space cooling consumes significant energy in India and is expected to increase in near future due to urbanization. Therefore, reducing energy consumption in buildings is becoming an interesting problem. However, energy profiles of space cooling depend on the thermal dynamics of the building, which in turn are influenced by numerous factors such as solar radiation, occupancy, inter-zone effects, and other factors which are themselves time varying. Modelling the effects of these aspects on building thermal dynamics is essential for reducing the energy consumption. Within the building control, model predictive control (MPC) is a promising technique due to its ability to include climate, occupancy, and other disturbance parameter to compute the optimal cooling energy to be supplied to a building [19, 20]. However, designing MPC requires building thermal models capturing the effect of disturbances [2]. Therefore, disturbance observers for buildings become more important.

Two approaches for modelling buildings are forward and data driven [10]. The actual building properties which include systems geometry, orientation and other aspects are modelled in software to predict the energy performance in forward modelling. This is a detailed information model involving material type, architecture, fenestration and other aspects. Usually, developing such models for energy analysis is very difficult. Data-driven approaches use data to build a mathematical model and are useful for designing controllers. In practice, influence of the disturbance parameters is usually estimated from empirical formulas, for example Air-Conditioning, Heating, and Refrigeration Institute formula. An empirical method which formalizes cooling load computation as a software was presented in ref. [1]. Regression based analysis for predicting disturbance effects was presented in ref. [18] to compute the cooling load using off-line data. The use of artificial neural network to model disturbance was studied in refs. [4, 5, 9, 10, 17]. Data-mining techniques such as support vector machines (SVR), artificial neural networks, classification and regression tree, chi-squared automatic interaction detection, general linear regression, and ensemble inference model were studied for modelling cooling loads.

in ref. [3]. Similarly, the investigation in ref. [6] used support vector machines (SVM) to predict hourly stray disturbances in the buildings.

Dynamical simulation tools such as DOE-2, EnergyPlus, TRNSYS, BDA, ESP-r, BLAST, IDA, etc. [16] are being widely studied in the literature for modelling building thermal dynamics. The use of models to estimate cooling loads with two software Building Controls Virtual Test Bed (BCTVB) and EnergyPlus was studied

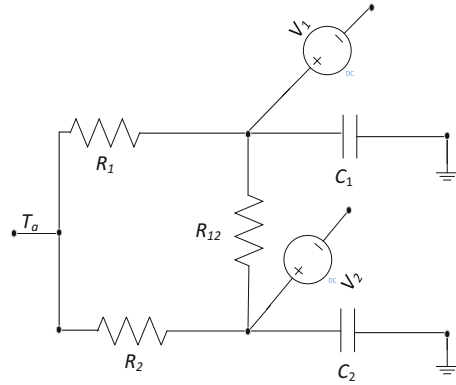
in ref. [7]. The estimation of thermal load in buildings by including factors such as construction, geometry, layout, climate, etc. was proposed in ref. [16]. In addition, the proposed method studied the use of artificial neural network to capture the cooling load in buildings. The investigation in refs. [11, 12] presented a model-based approach for estimating the cooling load in the buildings. The proposed approach used a reduced order state space thermodynamic model and extended Kalman filter for estimating the cooling loads. The advantage of the method is the development of soft-sensor which can be used to measure the cooling load dynamically. However, the investigation had a restrictive assumption that the system is linearized around an operating point to design the extended Kalman filter (EKF). Considering the multiple factors influencing the building thermal dynamics and complex behaviours this may not be always possible. Further, it requires linearization during each computing epoch which makes their adoption in building automation systems complex. A stable observer is designed for disturbance estimation in ref. [13], and disturbance decoupling and regulation are done by feedback linearization with back-stepping methods. The closed-loop stability of the system is assured for measurable disturbances. Since the disturbances are not always measurable, the dynamics of the observer influence the closed-loop stability of the system. A method to maintain temperature bounds within user-defined comfort bands using these methods reveals the lack of techniques for adaptive model identification for building thermal disturbance predictions.

This chapter proposes an adaptive disturbance observer for non-linear and time-varying disturbance in the building thermal dynamics. The observer has two blocks: (i) digital equivalent of the building thermal model, and (ii) gain adaptation algorithm. The digital equivalent is a thermal model of the building modelled as a circuit model capturing the effects of various influences. The gain adaptation algorithm is a simple gradient-descent approach to adjust the gain to reduce the estimation error. Due to the proposed approach, a disturbance observer which estimates the states is designed. The main contributions of this chapter are an adaptive disturbance estimator and derive the conditions for which the estimator works. This chapter is organized into five sections. Preliminaries and problem formulation are presented in Sect. 2. Disturbance observer design is discussed in Sect. 3 and the simulation results are presented in Sect. 4. Conclusions and future course of investigation are presented in Sect. 5.

18.2 Preliminaries

There are numerous factors that influence the building energy consumption which include ambient temperature, humidity, stray heating in the room, cooling loads and solar radiation into the rooms. These factors are non-linear and time-varying thereby making thermal dynamics of buildings complex and difficult to estimate the effect of disturbances. Therefore, to understand the building thermal dynamics, it is

Fig. 18.1 Equivalent circuit of the thermal zone



imperative to design a disturbance observer. Consider a building with two rooms to be controlled using a thermostat whose turn-on and turn-off time could be varied using an optimization-based control as in refs. [14, 15]. Following ref. [8], the building can be modelled using an equivalent circuit shown in Fig. 18.1. The voltage sources V_1 and V_2 at the nodes indicate the disturbance inputs and the interactions between the temperatures are modelled using the resistance R_{12} . In addition, the disturbance T_a models the effects of external heating on the building due to ambient temperature, solar radiation, etc. Thermal resistance of the zones is modelled using R_1 and R_2 , respectively. Similarly, thermal capacitance is given by C_1 and C_2 . The values can be computed from direct simulation in EnergyPlus.

18.2.1 Notation and Definitions

i, j	index of rooms
$U A_i$	overall heat transfer coefficient between room i and j in $\text{kW } ^\circ\text{C}^{-1}$ of zone i
T_a	ambient temperature incident on the building in $^\circ\text{C}$
T_i	temperature of the room i in $^\circ\text{C}$, R_i —thermal resistance of room i in $\text{kW } ^\circ\text{C}^{-1}$
C_i	thermal capacitance of room i in $\text{J } ^\circ\text{C}^{-1}$, V_i —cooling inputs to the rooms in kW

18.2.2 Thermal Dynamics of the Air Conditioner

Consider that the state of the room with the air conditioner is described by $x_1 = T_1$ and that of the adjacent room is given by $x_2 = T_2$. In addition, let W_1 and W_2 denote the heat gains in the zone 1 and zone 2 due to ambient conditions and internal heat gains. The heat gain due to solar radiation is given by W_s and the T_a and T_o denote the heating due to ambient conditions and occupancy respectively. Then the system dynamics is given by

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + Ew(t), \\ y(t) &= Cx(t),\end{aligned}\tag{18.1}$$

where $x(t) \in \mathbb{R}^2$ denotes the zone temperatures $T_1(t)$ and $T_2(t)$, the cooling inputs are given by $u(t) \in \mathbb{R}^2$ which model the duration for which the air conditioner is turned ON/OFF in each rooms, $w(t) \in \mathbb{R}^{4 \times 2}$ models the internal heat gains in rooms, solar heat gain, heat gain due to ambient temperature, and occupancy, respectively. The sensor measurements are given by $y(t)$. The matrices A , B , E , and C (observation) matrices are given by

$$\begin{aligned}A &= \begin{bmatrix} -\left(\frac{1}{R_1 C_1} + \frac{1}{R_1 C_1}\right) & \frac{1}{R_{12} C_1} \\ \frac{1}{R_{12} C_1} & -\left(\frac{1}{R_{12} C_1} + \frac{1}{R_2 C_2}\right) \end{bmatrix}, & B &= \begin{bmatrix} -\left(\frac{1}{C_1}\right) & 0 \\ 0 & -\left(\frac{1}{C_1}\right) \end{bmatrix}, \\ E &= \begin{bmatrix} -\left(\frac{1}{C_1}\right) & 0 & -\frac{\alpha_1}{C_1} & -\frac{1}{R_1 C_1} \\ 0 & -\left(\frac{1}{C_1}\right) & -\frac{\alpha_2}{C_2} & -\frac{1}{R_2 C_2} \end{bmatrix}, & C &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned}$$

18.2.3 Problem Definition

The disturbances acting on the system are not only time varying but also depend on the system parameters which are computed from simulations or knowledge of the physics.

The need to implement such an observer in a digital computer compliments the complexity further. Usually exact computation of the parameters is not feasible and this introduces modelling errors. In addition, the disturbances are not completely observed, while inferences can be made on their values through the state, not exact measurements of the disturbances are possible. Therefore, an observer for both state and disturbances are required. However, designing such an observer requires certain assumptions which are defined here. The discrete-time representation for a sampling time δ of the (18.1) is given by

$$\begin{aligned}x(k+1) &= \Phi x(k) + \Gamma u(k) + Gw(k) \\ y(k) &= Cx(k)\end{aligned}\tag{18.2}$$

(i) The matrix E has full column rank and the disturbances enter the system as additive disturbance. (ii) There exists a sampling time δ for which the disturbance inputs vary slowly and, in addition, preserve the column rank of G , and μ is a small real constant

$$\|w(k+1) - w(k)\|_2 \leq \mu. \quad (18.3)$$

(iii) The matrix \emptyset has full rank and its eigenvalues are located inside the unit circle.

Now considering the disturbance to be the states of the system and by defining the augmented state as $z = [x \ w]^T$, we have

$$\begin{aligned} z(k+1) &= \phi z(k) + \Gamma_1 u(k), \\ y(k) &= [C \ 0]z(k), \end{aligned} \quad (18.4)$$

where $\phi = \begin{bmatrix} \emptyset & 0 \\ 0 & G \end{bmatrix}$, $\Gamma_1 = \begin{bmatrix} \Gamma \\ 0 \end{bmatrix}$, and 0 denotes the vector of zeros commensurate with the dimensions of the disturbance vector.

Problem 1 The problem is to design a disturbance and state observer for the system in Eq. (18.1) on single-board computers using the formulation proposed in Eq. (18.4) and with assumption 1–3 defined above. However, the disturbance vectors are changing with time and a static observer may not suffice. Therefore, we propose to design an adaptive disturbance observer for the system.

Problem 2 Suppose that there exists an adaptive disturbance estimator for the time-varying disturbances. Is it possible to formulate the conditions for which the estimator works? The answers to the above questions are considered in the following sections.

18.3 Adaptive Disturbance Estimator

Disturbance observers can be constructed considering the augmented state to be the state of the system and disturbance inputs. The block diagram of the adaptive disturbance observer is shown in Fig. 18.2. It consists of the conventional observer from which the disturbance estimator works. With a bounded control input one can observe that the disturbance and state can reduce but there are no guarantees of the disturbance going to zero. Therefore, we define the error term

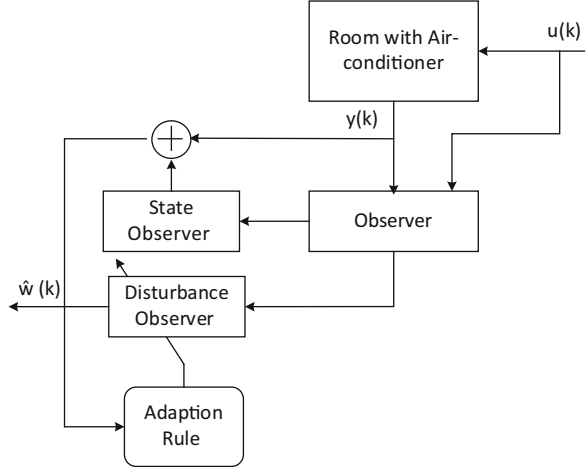
$$\varepsilon(k) = y(k) - C\hat{x}(k). \quad (18.5)$$

To adjust the disturbance observer, we use a simple gradient descent rule given by

$$L_w(k+1) = -\gamma \times \varepsilon(k) \frac{\partial \varepsilon(k)}{\partial L_w} \quad (18.6)$$

with the adjusted gain the closed loop dynamics of the system is given by

Fig. 18.2 Disturbance observer for air conditioners



$$\begin{aligned} x(k+1) &= (\emptyset - \Gamma L)x(k) + (G - \Gamma L_w)w(k) + u(k), \\ w(k+1) &= Gw(k) - L_w(y(k) - C\hat{x}(k)), \end{aligned} \quad (18.7)$$

where $u(k) = -K\tilde{x}(k) - K_w \tilde{w}(k)$ with \tilde{x} and \tilde{w} denoting the closed loop states of the system with the observers. These are static gains that can be obtained from the assumptions (i)–(iii).

Conversely, the gain $L_w(k)$ determines whether the disturbance goes to zero or not. Due to assumption (ii), we can see that the slow-moving dynamics will decay the disturbance to zero.

18.3.1 Conditions for Observability of Disturbance Estimator

Remark Suppose that the gain of the observer is given by L , the closed-loop dynamics is given by

$$\begin{aligned} x(k+1) &= (\emptyset - \Gamma L)x(k) + (G - \Gamma L_w)w(k) + u(k), \\ w(k+1) &= Gw(k) - L_w(y(k) - C\hat{x}(k)). \end{aligned} \quad (18.8)$$

The assumption (ii) signifies that the disturbance cannot change faster between sampling interval and we denote the change in disturbance as δ_w and is given by

$$\delta_w = \emptyset_w \times w \quad (18.9)$$

One can see that due to the fact that the disturbance cannot change significantly within a given interval, the $\|\delta_w\|_{\ell_2}$ is upper bounded by δ_μ . We define the following performance metric for disturbance

$$I = \sum_{\tau=t}^N \delta_w(\tau)^T Q \delta_w(\tau) \forall \tau \in \{t, t+1, \dots, t+N\}, \quad (18.10)$$

where Q is a positive definite matrix. The metric I is convex in the domain $[\delta_w(k), \delta(k) * \mu]$. The observability of the disturbance for a given time period can be computed from the following measure:

$$W_0 = \sum_{\tau=t}^N (G^T C_w^T C_w G)^\tau, \quad (18.11)$$

$$W_0 - \sum_{\tau=t}^N G^T W_0 G \leq C_w^T C_w, \quad (18.12)$$

where $w = C_w v$, another vector v from which the disturbance vector can be observed. The necessary and sufficient condition for the observability of the disturbance vector is then given by Eq. (18.13) and generalizing the result to N Steps, we have Eq. (18.14):

$$W_0 - \sum_{\tau=t}^{t+1} G^T W_0 G \leq \delta(k) * \mu, \quad (18.13)$$

$$W_0 - \sum_{\tau=t}^{t+1} G^T W_0 G \leq N \times \delta(k) * \mu \quad (18.14)$$

One can observe that

$$i \leq N \times \delta(k) * \mu \quad (18.15)$$

18.4 Results

This section presents the simulation results of the disturbance observer for building air-conditioner system described in Sect. 18.2. We assume a sampling time of 15 s and the heat transfer coefficient of 76.1 was used in our computations. The capacitances of the zones were computed to be 13,500 and that of the adjacent room was around 5000. The simulations were performed with the air conditioner being at an initial temperature of 24 °C, ambient temperature varies to about 32 C, heating due to solar radiation and occupancy being limited to 6 °C, and stray heating is less than 1 °C. The estimates of the state and disturbance for a period of 2 days are shown in Fig. 18.3.

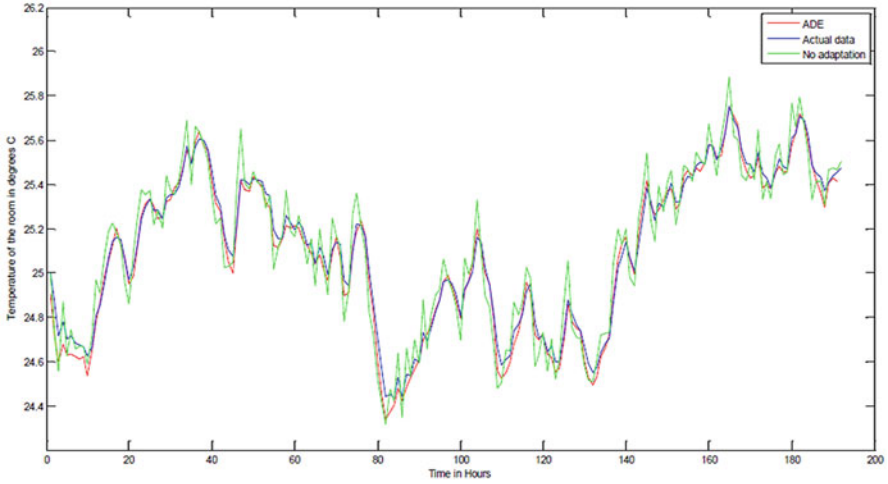


Fig. 18.3 State estimation with and without adaptive disturbance observer

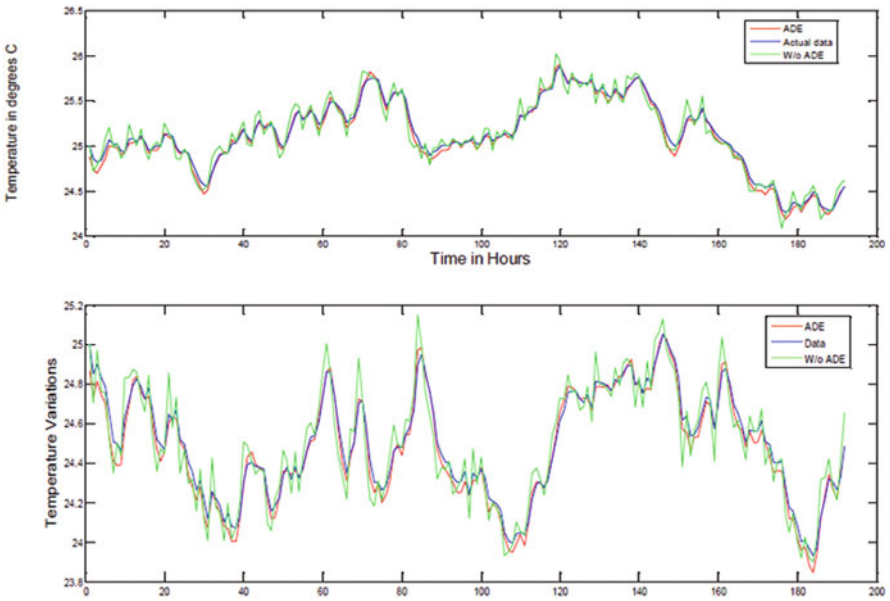


Fig. 18.4 State estimation with and without adaptive disturbance observer for changes in occupancy and ambient temperature

The variations in the temperature for changes in occupancy and ambient temperature introduced in simulations are shown in Fig. 18.4. One can see that adaptive disturbance estimator (ADE) performs well for the disturbances than conventional observer. The ADE closely captures the variation and reduces the disturbances.

18.5 Conclusions

This chapter presented an adaptive disturbance estimator for measuring stray heating and disturbances in buildings cooled using air conditioners. In addition, it provided the conditions on which the disturbance observer works. The proposed observer was demonstrated on a building data equipped with air conditioners using simulations. Our results showed that the adaptive disturbance estimator provided better estimates of the states due to suppression of disturbances and therefore, could be used with model-based control techniques for computing the disturbance estimates. Formulating the disturbance estimator as an optimization model and developing numerical methods to solve the problem are future course of this investigation.

References

1. Ansari FA, Mokhtar AS, Abbas KA, Adam NM (2005) A simple approach for building cooling load estimation. *Am J Environ Sci* 1(3):209–212
2. Arguello-Serrano B, Velez-Reyes M (1999) Nonlinear control of a heating, ventilating, and air conditioning system with thermal load estimation. *IEEE Trans Control Syst Technol* 7(1):56–63
3. Chou J-S, Bui D-K (2014) Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy Buildings* 82:437–446
4. Gambino G, Verrilli F, Canelli M, Russo A, Himanka M, Sasso M, Srinivasan S, Del Vecchio C, Glielmo L (2016) Optimal operation of a district heating power plant with thermal energy storage. In: American control conference (ACC), 2016, IEEE, pp 2334–2339
5. Hou Z, Lian Z, Yao Y, Yuan X (2006) Cooling-load prediction by the combination of rough set theory and an artificial neural-network based on data-fusion technique. *Appl Energy* 83(9):1033–1046
6. Li Q, Meng Q, Cai J, Yoshino H, Mochida A (2009) Applying support vector machine to predict hourly cooling load in the building. *Appl Energy* 86(10):2249–2256
7. Li X, Wen J (2014) Building energy consumption on-line forecasting using physics based system identification. *Energy Buildings* 82:1–12
8. Ma Z, Wang S (2009) Energy efficient control of variable speed pumps in complex building central air-conditioning systems. *Energy Buildings* 41(2):197–205
9. Sara MC M^a, Leal V^ı MS, Horta IM (2017) Modelling the relationship between heating energy use and indoor temperatures in residential buildings through artificial neural networks considering occupant behavior. *Energy Buildings* 151:332–343
10. Nilashi M, Dalvi-Esfahani M, Ibrahim O, Bagherifard K, Mardani A, Zakuan N (2017) A soft computing method for the prediction of energy performance of residential buildings. *Measurement* 109:268–280
11. O’Neill Z, Narayanan S, Brahme R (2010) Modelbased thermal load estimation in buildings. *Proc SimBuild* 4(1):474–481
12. O’Neill Z, Narayanan S, Brahme R (2010) Modelbased thermal load estimation in buildings. *Proc SimBuild* 4(1):474–481
13. Semsar-Kazerouni E, Yazdanpanah MJ, Lucas C (2008) Nonlinear control and disturbance decoupling of hvac systems using feedback linearization and backstepping with load estimation. *IEEE Trans Control Syst Technol* 16(5):918–929
14. Soudari M, Kaparin V, Srinivasan S, Seshadhri S, Kotta U (2018) Predictive smart thermostat controller for heating, ventilation, and air-conditioning systems. *Proc Est Acad Sci* 67(3):291–299

15. Soudari M, Srinivasan S, Balasubramanian S, Vain J, Kotta U (2016) Learning based personalized energy management systems for residential buildings. *Energy Buildings* 127:953–968
16. Turhan C, Kazanasmaz T, Uygun IE, Ekmen KE, Akkurt GG (2014) Comparative study of a building energy performance software (kep-iyte-ess) and ann-based building heat load estimation. *Energy Buildings* 85:115–125
17. Verrilli F, Srinivasan S, Gambino G, Canelli M, Himanka M (2017) Carmen Del Vecchio, Maurizio Sasso, and Luigi Glielmo. Model predictive control-based optimal operations of district heating system with thermal energy storage and flexible loads. *IEEE Trans Autom Sci Eng* 14(2):547–557
18. Yao Y, Lian Z, Liu S, Hou Z (2004) Hourly cooling load prediction by a combined forecasting model based on analytic hierarchy process. *Int J Therm Sci* 43(11):1107–1118
19. Gambino G, Verrilli F, Canelli M, Russo A, Himanka M, Sasso M, . . . , Glielmo L (2016, July) Optimal operation of a district heating power plant with thermal energy storage. In *American control conference (ACC)*, 2016, IEEE, pp 2334–2339
20. Belikov J, Kotta Ü, Srinivasan S, Kaldmäe A, Halturina K (2013) On exact feedback linearization of HVAC systems. In *Process control (PC)*, 2013 International Conference on, pp 353–358, IEEE

Chapter 19

HTSA Optimized PID-Based MPPT for Solar PV System



Shashikant and Binod Shaw

Abstract PV system is generally applied with DC–DC converter and a battery. The system is always operated at maximum power point, and its tracking is known as maximum power point tracking (MPPT). Several comparisons have been based on MPPT. In this article, a comparison has been made with conventional perturb and observe (P&O) with intelligent fuzzy logic controller. In this article, the system is designed to operate for an islanding mode without a battery which would reduce the cost of a battery. The model is simulated in MATLAB and the obtained simulated result shows that the performance of intelligent optimized fuzzy logic controller is better than conventional P&O method.

Keywords Proportional integral and derivative (PID) · Heat transfer search algorithm (HTSA) · Fuzzy proportional integral and derivative (FPID) · Perturb and observe (P&O)

Abbreviations

MPPT	maximum power point tracking
P&O	perturb and observe.
PID	proportional integrator and derivative.
FPID	fuzzy proportional integrator and derivative.
HTSA	heat transfer search algorithm.
CDF	conduction factor.
COF	convection factor.
RDF	radiation factor.
TCF	temperature change factor.

Shashikant (✉) · B. Shaw
National Institute of Technology, Raipur, Chhattisgarh, India

19.1 Introduction

Renewable energy resources such as solar and wind are gaining popularity in recent years. It is environment friendly and has no fuel cost. The fuel for photovoltaic system is sun's energy and fuel for wind energy conversion system is wind, both are freely available natural resources. The characteristics of both sources are highly non-linear. The characteristics of photovoltaic system depend on solar irradiation and temperature. So, a control algorithm is needed to get maximum power from photovoltaic system, such a concept is known as MPPT (maximum power point tracking). Hence, every solar cell is operated at the point of maximum power.

Several conventional methods are developed to track maximum power point such as incremental conductance, linear current control-based MPPT, temperature-based MPPT, constant voltage-based method, perturb and observe (P&O) method, neural network-based MPPT, etc. MPPT algorithm needs to be improved because operating point of PV panel oscillates around maximum power point, that is unstable MPP, which is illustrated in Ref. [1], for tracking MPPs at a high speed with accuracy, a DC–DC converter such as cuk converter is used with incremental conductance algorithm illustrated in Ref. [2]. By comparing the incremental conductance algorithm with P&O method under fast changing of solar irradiation, it is found that the incremental conductance shows better performance over P&O which is illustrated in Ref. [3]. Apart from conventional MPPTs a new approach, that is intelligent fuzzy MPPT controllers under various climatic conditions have been studied and a literature is provided which tells that performance of PV system in terms of tracking in speed, accuracy and stability is more advanced than the conventional MPPT methods which are well illustrated in Ref. [4].

This chapter presents an optimized fuzzy logic controlled-based MPPT algorithm using boost converter. Boost converter is used because the output voltage of module is low for practical application. So, for practical application the voltage needs to be boosted, hence a DC–DC boost converter is presented in this paper. Since the boost converter is operating at continuous current mode to extract maximum power from solar cells [5], efficiency of the system is improved. A heat transfer search algorithm is applied to obtain the gain parameters of proportional integrator and derivative (PID) controller and the range of controller is defined by the output of fuzzy. The duty cycle of boost converter is varied based on the load matching mechanism, to achieve constant power and voltage for desired application. Comparison has been made between conventional P&O and optimized fuzzy PID controller-based MPPT algorithm.

19.2 System Investigation

The system comprises of PV panel, MPPT algorithm, boost converter and controller as portrayed in Fig. 19.1. Solar irradiation and temperature are the two major factors which affect the efficiency of solar cell. Therefore, these two factors are taken into

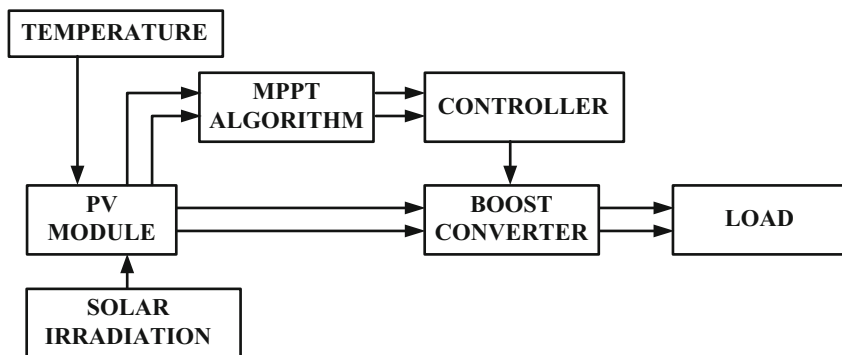
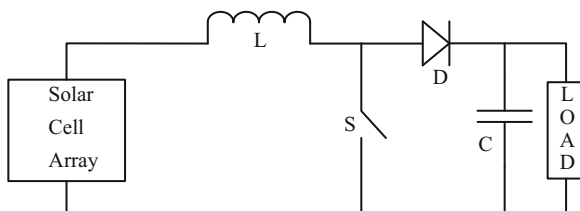


Fig. 19.1 Block diagram of PV system

Fig. 19.2 Boost converter model



consideration, and the variations in solar irradiation with temperature at 45 °C are taken at different time intervals to observe the results.

19.2.1 Boost Converter

The basic operation of boost converter is to boost the operating quantity. The output voltage of PV system is generally small; in order to increase the voltage rating of the system, multiple solar arrays are connected in series or in parallel. Another way to increase the voltage rating of PV system is using boost or buck–boost converter. Buck–boost converter is used when the input voltage is greater than or less than the output voltage. Boost converter is used when the input voltage is always less than the output voltage. In this chapter, the system is designed for low power application, hence boost converter is taken into consideration. Boost converter model is presented in Fig. 19.2, in which L represents inductor, S is switch, D is diode and C is capacitor.

19.3 Controllers

19.3.1 PID Controller

PID refers to proportional, integral, and derivative controller. P mode in PID controller eliminates error oscillation, D mode processes the speed of controller, that is it gives fast reaction with respect to change in signal input and integral mode fuels the control signal to attain error towards zero. These overall dynamics are necessary and important for industrial applications. Figure 19.3 represents the block diagram of PID controller, in which K_p represents the proportional constant, K_i represents the integral constant, K_d represents the proportional constant, $e(t)$ represents the error signal obtained from the difference of processed value and reference value and finally the $u(t)$ represents the controlled output of the controller. PID controller is expressed mathematically as Eq. (19.1) in time domain and Eq. (19.2) in Laplace domain:

$$u(t) = K_p \cdot e(t) + K_i \int_0^t e(t) \cdot dt + K_d \frac{d}{dt} e(t) \tag{19.1}$$

$$U(s) = (K_p K_P + K_i K_I * (1/S) + K_d K_D * S) * E(S) \tag{19.2}$$

19.4 Maximum Power Point Tracking (MPPT)

The duty cycle of boost converter is varied based on the MPPT algorithm applied as shown in Figs. 19.4 and 19.5 to meet the load requirements. The cost of battery is reduced since the boost converter is able to meet the load requirements. In the chapter, change in irradiation is taken from 1000 to 800 W/m² just to differentiate the observed results, but in practical change of irradiation is less and based on the calculation of V_{ref} the boost converter output is maintained constant which is

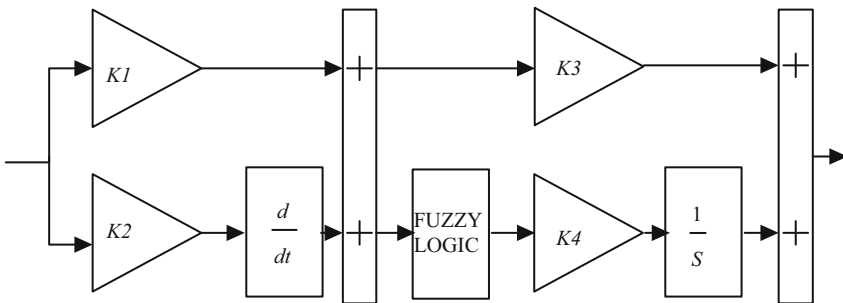


Fig. 19.3 Block diagram of FPID controller

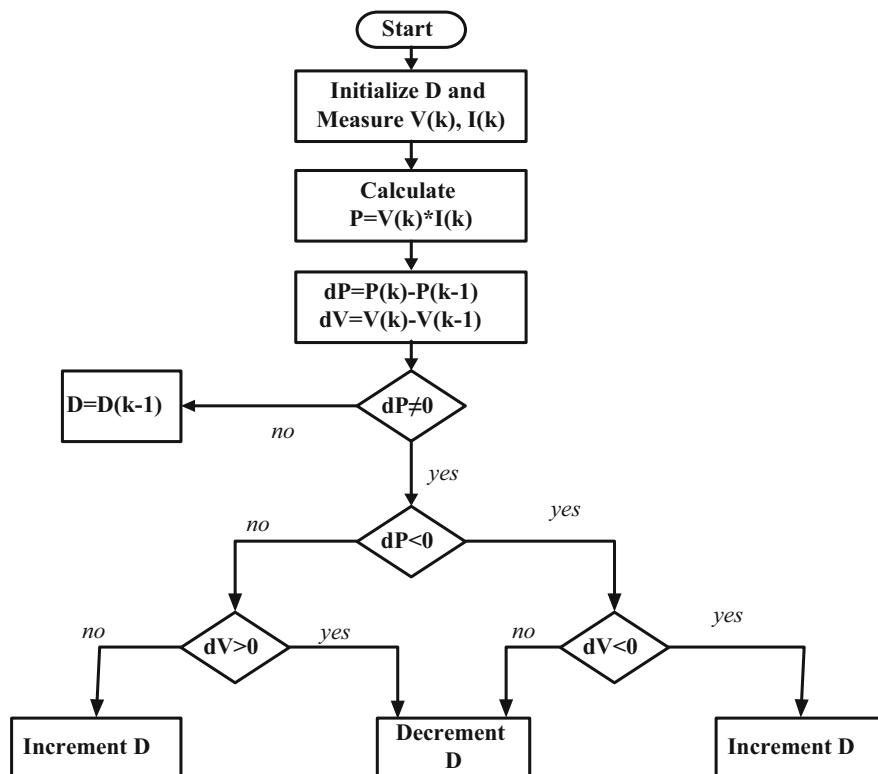


Fig. 19.4 MPPT algorithm for P&O

desirable for load. The boost converter parameters are taken from Ref. [6]. The concept of intelligent fuzzy controller is taken from Ref. [7].

19.5 Heat Transfer Search Algorithm (HTSA)

HTSA was proposed by Savasani and Patel in 2005 [8]. ‘This algorithm is based on the natural law of thermodynamics and the principle of thermal equilibrium of the system’ [8]. There are three modes of heat transfer such as conduction, convection and radiation. Thermodynamically imbalanced systems always try to attain thermal equilibrium state. Therefore, in HTSA radiation, convection and conduction try to reach an equilibrium state. Following are the steps involved in HTSA to obtain optimum value or tuned value of controller:

1. Initialize the population N and design variable D , which gives initial position ‘ H ’ of dimension $[N \times D]$.
2. Initialize conduction factor (CDF), convection factor (COF), radiation factor (RDF) and Itermax.

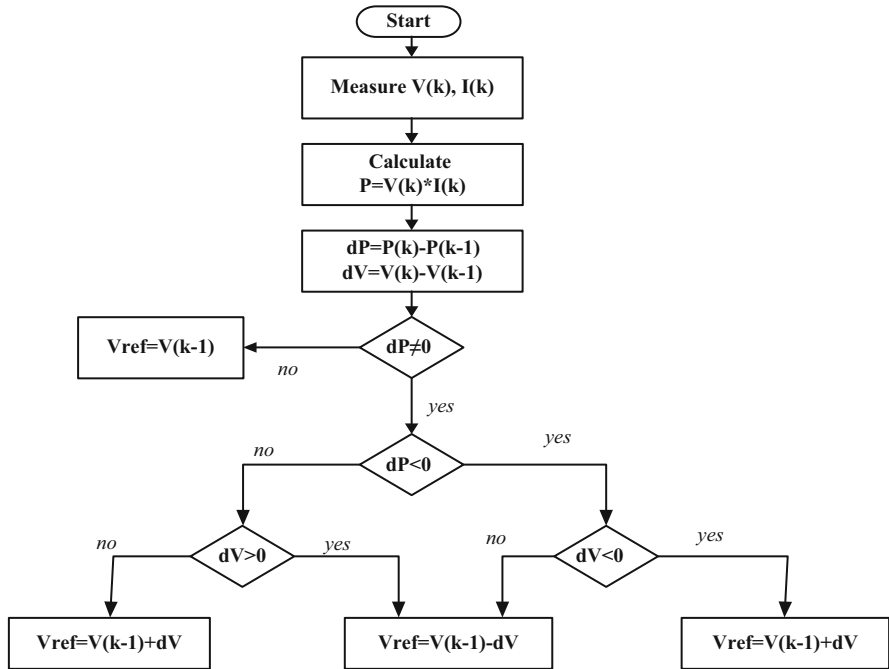


Fig. 19.5 MPPT algorithm for PID and FPID

3. Evaluate the fitness function.
4. Now, randomly generate a number R whose value lies between 0 and 1, through which conduction, convection and radiation process follows for updated positions.

Conduction: $R < 0.33$		
If $F \leq C$	$X_{j,i}^{new} = X_{k,i}^{old} - R^2 * X_{k,i}^{old}$	If $f(X_j) > f(X_k)$
	$X_{k,i}^{new} = X_{j,i}^{old} - R^2 * X_{j,i}^{old}$	If $f(X_k) > f(X_j)$
If $F \geq C$	$X_{j,i}^{new} = X_{k,i}^{old} - r_i * X_{k,i}^{old}$	If $f(X_j) > f(X_k)$
	$X_{k,i}^{new} = X_{j,i}^{old} - r_i * X_{j,i}^{old}$	If $f(X_k) > f(X_j)$
Radiation: $R > 0.66$		
If $F \leq C$	$X_{j,i}^{new} = X_{j,i}^{old} + R * (X_{k,i}^{old} - X_{j,i}^{old})$	If $f(X_j) > f(X_k)$
	$X_{j,i}^{new} = X_{j,i}^{old} + R * (X_{j,i}^{old} - X_{k,i}^{old})$	If $f(X_k) > f(X_j)$
If $F \geq C$	$X_{j,i}^{new} = X_{j,i}^{old} + r_i * (X_{k,i}^{old} - X_{j,i}^{old})$	If $f(X_j) > f(X_k)$
	$X_{j,i}^{new} = X_{j,i}^{old} + r_i * (X_{j,i}^{old} - X_{k,i}^{old})$	If $f(X_k) > f(X_j)$
Convection: $0.33 < R < 0.66$		
If $F \leq C$	$X_{j,i}^{new} = X_{j,i}^{old} + R * (X_s - X_{ms} * TCF)$	
	$TCF = \text{abs}(R - r_i)$	$g \leq G_{max}/COF$
If $F \geq C$	$TCF = \text{round}(1 + r_i)$	$g \geq G_{max}/COF$

where F is the fitness value or generation (g), r_i is a random number between 0 and 1, $j = 1, 2, \dots, n$, $j \neq k$, $k \in (1, 2, \dots, n)$, from the population k is a randomly selected solution, $i \in (1, 2, \dots, m)$, which is a randomly selected design variable; $X_{k,i}$ and $X_{j,i}$ correspond to the temperature gradient, X_s and X_{ms} are the surrounding temperature and mean temperature of the system, $C = F_{\max}/\text{CDF}$ for conduction and for radiation phase $C = F_{\max}/\text{RDF}$, the value of conduction factor (CDF) and radiation factor (RDF) is taken as 2. TCF is the temperature change factor which balances the exploration and exploitation.

5. Updated positions are stored in variable X and its calculated fitness value is stored in a new fitness variable F_{new} .
6. Upon comparison of fitness value, that is F_{new} and F_{old} (F). Best fitness value is obtained that contains optimum value of design variables.
7. To get more optimum values repeat the steps from 3 to 6.

The flow chart of HTSA is presented in Fig. 19.6; HTSA involves three ways to update the position of population to determine the optimal operating point, that is conduction, convection and radiation. Like other optimization techniques, a weight factor determines the movement of variables, similarly in HTSA, ' R ' acts as a weight factor and the special feature of HTSA as it undergoes three processes to update the position which gives result closer to the best optimal point.

19.6 Result and Discussion

The system is simulated in MATLAB as per the block diagram portrayed in Fig. 19.1. The value of L , C and R is taken as 100 mH, 3300 μf and 5 Ω , respectively for boost converter. For the same boost converter, the comparison is made among P&O, optimized PID and fuzzy proportional integrator and derivative (FPID) for calculation of MPPT. Conventional P&O uses calculation of MPPT based on duty ratio and for PID and FPID both are applied to the same algorithm to calculate $V_{\text{reference}}$. For analysis purpose the change in irradiation is taken in simulation at time $t = 0.3$ s, the irradiation is changed from 1000 to 800 W/m^2 .

Figure 19.7a–c represents the power, voltage and current profile, respectively. It is observed that as the solar irradiation is increased the output power increases or vice versa. The amount of power extracted from solar with boost converter in case of P&O is around 30 VA. The power can be extracted more by applying optimized PID and FPID with change in MPPT algorithm. The power voltage and current available at load are tabulated in Table 19.1. The obtained value of FPID controllers k_1 , k_2 , k_3 and k_4 is 0.7615, 1.1361, 0.1526 and 0.1088, respectively followed from Fig. 19.2 and the values for conventional PID controllers are 0.0711, 0.9079 and 1.1260.

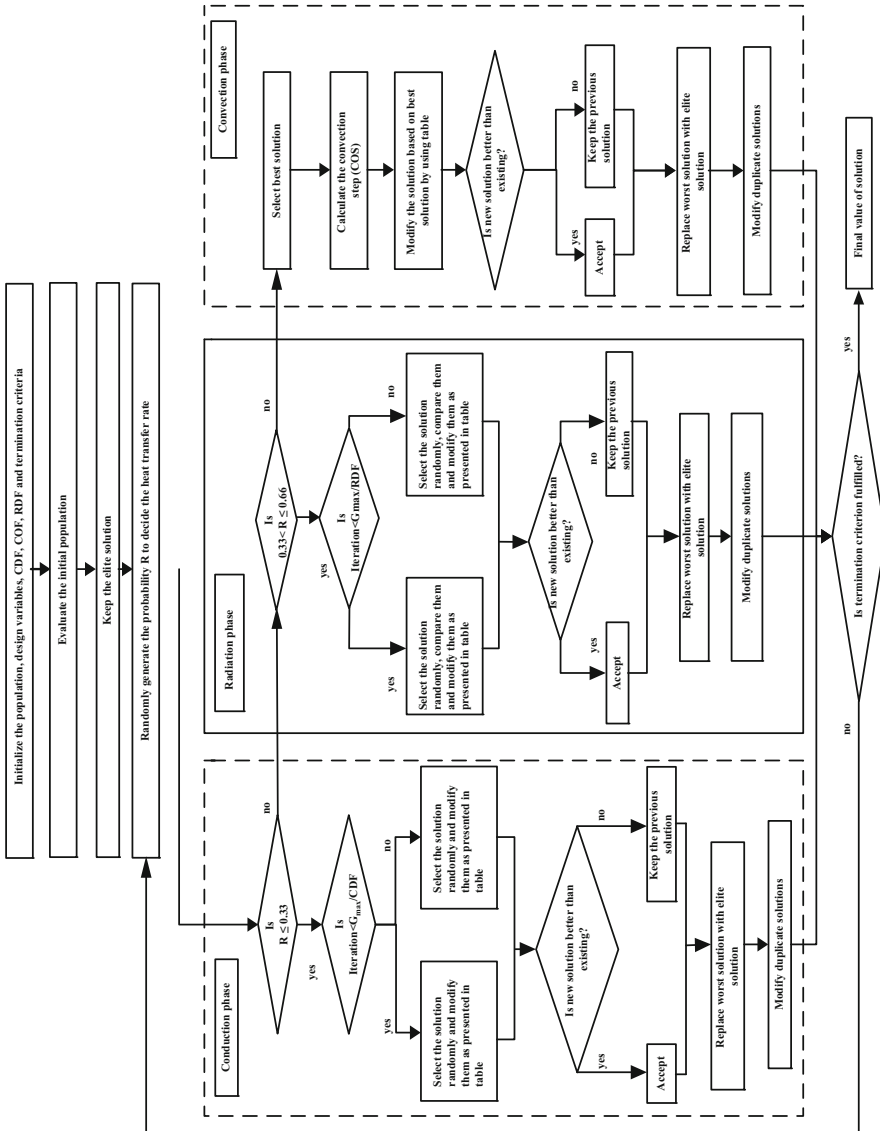


Fig. 19.6 Flow chart of heat transfer search algorithm

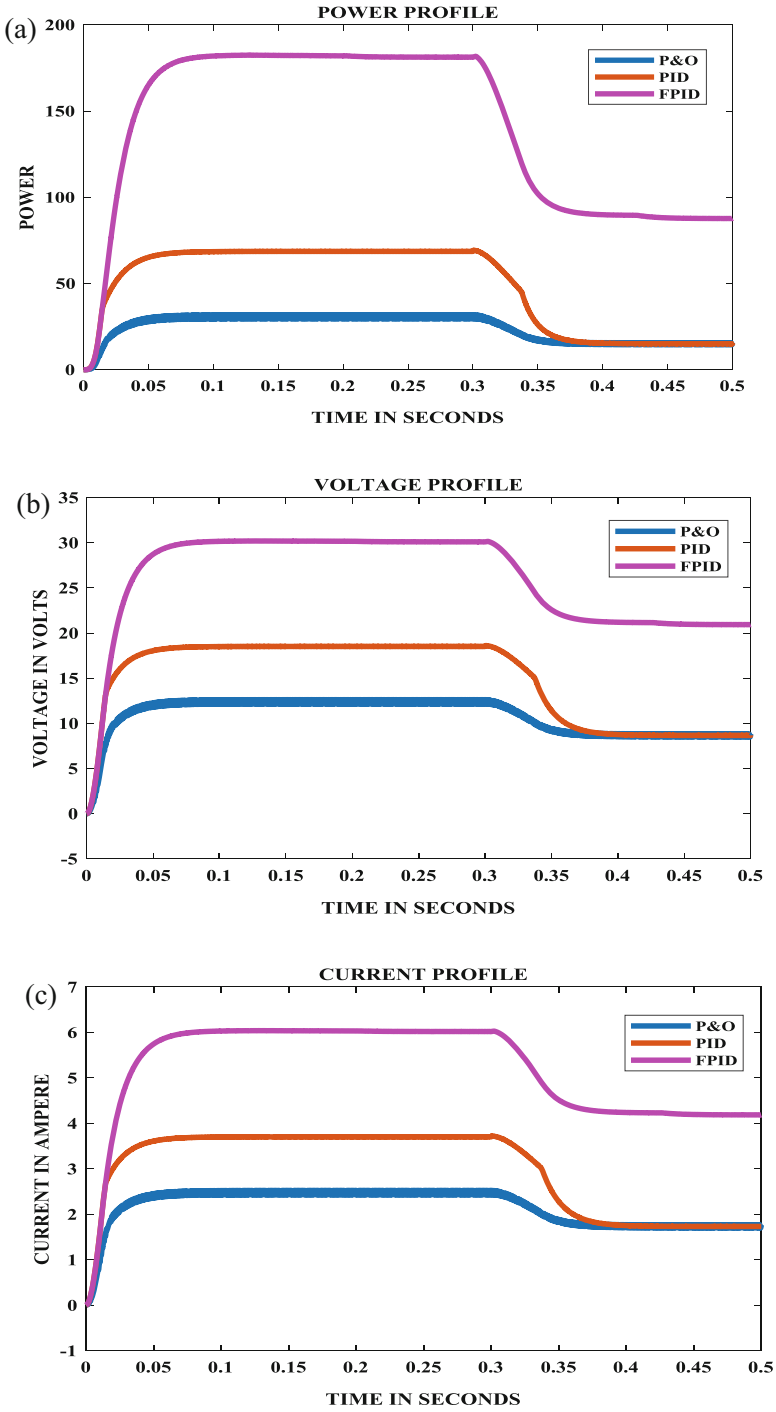


Fig. 19.7 (a) Power profile for boost converter, (b) voltage profile for boost converter and (c) current profile for boost converter

Table 19.1 Output parameters of boost converter

Method	Parameters irradiation 1000 W/m ²			Parameters irradiation 800 W/m ²		
	Voltage	Current	Power	Voltage	Current	Power
P&O	12	2.4	30	8	1.7	15.7
PID	18	3	68	8	1.7	15.7
FPID	30	6	182	20	4	88.4

19.7 Conclusion

A boost converter-based solar PV system is presented in this article. A conventional P&O-based MPPT algorithm was compared with FPID and conventional PID-based MPPT algorithm. It is observed that the FPID and conventional PID-based MPPT algorithm yields better result over P&O-based MPPT algorithm. Due to continuous conduction of boost converter the efficiency of the system is increased. The result presented in this article is analysed only for isolated loads. Hence, it concludes that the intelligent fuzzy controller is well suited for solar PV system.

References

1. Sharma DK, Purohit G (2012) Advanced perturbation and observation (P&O) based maximum power point tracking (MPPT) of a solar photo-voltaic system. India International Conference on Power Electronics, IICPE, pp 3–7. <https://doi.org/10.1109/IICPE.2012.6450411>
2. Safari A, Mekhilef S (2011) Incremental conductance MPPT method for PV systems, pp 345–347
3. Kim T, Ahn H, Park S, Lee Y (2001) A novel maximum power point tracking control for photovoltaic, pp 0–3
4. Bendib B, Belmili H, Krim F (2015) A survey of the most used MPPT methods: conventional and advanced algorithms applied for photovoltaic systems. *Renew Sust Energ Rev* 45:637–648. <https://doi.org/10.1016/j.rser.2015.02.009>
5. Appelbaum J (1996) Advantage of boost vs. buck topology for maximum power (2), pp 355–358
6. Shaw B (2015) *Soft computing in data science*, vol 545. Springer, Singapore. <https://doi.org/10.1007/978-981-287-936-3>
7. Veerachary M, Senjyu T, Uezato K (2003) Neural-network-based maximum-power-point tracking of coupled-inductor interleaved-boost-converter-supplied PV system using fuzzy controller. *IEEE Trans Ind Electron* 50(4):749–758. <https://doi.org/10.1109/TIE.2003.814762>
8. Patel VK, Savsani VJ (2015) Heat transfer search (HTS): a novel optimization algorithm. *Inf Sci (Ny)* 324:217–246

Chapter 20

Performance Analysis of UFMC System with Different Prototype Filters for 5G Communication



M. Maheswari, N. R. Nagarajan, and M. Banupriya

Abstract The Universal Filtered Multi-Carrier (UFMC) system is a promising contender for next-generation 5G wireless communication. UFMC has better sub-carrier separation like Filter Bank Multi-Carrier (FBMC) and provides minimal complexity like Orthogonal Frequency Division Multiplexing (OFDM). Hence, in this chapter, the UFMC system has been proposed with different prototype filters for shaping the waveforms and to increase performance. The performance of the UFMC system is analysed for important parameters such as Power Spectral Density (PSD), Block Error Rate (BLER), Bit Error Rate (BER) and Peak-to-Average Power Ratio (PAPR). The filters considered for the analysis are the Hermite filter, the PHYDIAS filter, the Root Raised Cosine (RRC) filter and the Rectangular filter. The analysis shows that among the four filters, the Hermite filter offers the best performance. This filter provides 72%, 63%, 67% and 29% enhanced performance for the parameters PSD, BLER, BER and PAPR, respectively.

Keywords UFMC · OFDM · Prototype filters · Hermite · PHYDIAS filter · RRC filter · Rectangular filter

Abbreviations

BER	Bit Error Rate
BLER	Block Error Rate
FBMC	Filter Bank Multi-Carrier
FFT	Fast Fourier Transform
OFDM	Orthogonal Frequency Division Multiplexing
OOB	Out Of Band

M. Maheswari (✉) · N. R. Nagarajan
Department of ECE, K. Ramakrishnan College of Engineering, Trichy, Tamil Nadu, India

M. Banupriya
Department of ECE, Jayaram College of Engineering, Trichy, Tamil Nadu, India

PAPR	Peak to Average Power Ratio
PSD	Power Spectral Density
RRC	Root Raised Cosine
SLM	Selective Mapping
SNR	Signal-to-Noise Ratio
UFMC	Universal Filtered Multi-Carrier

20.1 Introduction

The growth in mobile traffic has accelerated significantly from 1G to the present 5G. This creates problems for the present mobile communication technology in supporting the anticipated growth in mobile traffic in the near future. To overcome this challenge, many researchers have started doing research in next-generation 5G communication systems [1]. In 4G LTE, the multi-carrier modulation Orthogonal Frequency Division Multiplexing (OFDM) technique has been used to increase data rate. There is no need to provide guard band for the sub-carrier and only a cyclic prefix is added to avoid inter-symbol interferences. The addition of the cyclic prefix causes a reduction in powerful phantom effectiveness. In addition, OFDM suffers from high PAPR and high power out of band (OOB). Because of these drawbacks, OFDM is not a suitable candidate for the next-generation mobile communication technology, 5G. 5G communication has applications with very high data rate connectivity of up to 10Gb/s. Hence, other multi-carrier modulation techniques such as Filter Bank Multi-Carrier (FBMC) and Universal Filtered Multi-Carrier (UFMC) are considered as alternative candidates [2].

In FBMC, each sub-carrier is filtered separately, unlike in OFDM, which uses a single filter for the entire sub-carrier [3]. This reduces OOB emission in FBMC. As the cyclic prefix is not added in FBMC, it has higher spectral efficiency than OFDM [4]. As FBMC uses separate filters for each sub-carrier, it increases the computational complexity and, therefore, it is also not suitable for 5G communication.

In UFMC, a separate filter is used for a group of sub-carriers and not for each sub-carrier, unlike in FBMC. Hence, computational complexity is reduced compared with FBMC which overcomes the dis-advantages of FBMC. This is the first research, to our knowledge, to have designed a UFMC system with the prototype filters Hermite, PHYDIAS, RRC and Rectangular. So far, only an FIR filter has been designed for UFMC.

20.2 Literature Survey

In Ref. [2], OFDM, FBMC and UFMC are compared and the Peak-to-Average Power Ratio (PAPR) of these modulation techniques is investigated. The authors [5] compare different modulation techniques such as FBMC, UFMC, Generalized Frequency-Division Multiplexing (GFDM) and resource-block filtered orthogonal frequency-division multiplexing with OFDM and propose FBMC shows better performance and insensitive to multi-user interfaces. The authors have explored [6] new derivative waveforms of OFDSM such as FBMC, UFMC and GFDM. Suitable spectral performance and computational complexity reduction have been investigated, which makes waveform selection for 5G more comfortable. In Ref. [7], the authors proposed a Hybrid PAPR reduction technique, Single-Carrier Universal Filtered Multicarrier (SC-UFMC), using Selective Mapping (SLM) and clipping. The performance of the proposed technique is evaluated for various design parameters, which showed that the hybrid technique provides better PAPR reduction than conventional SLM and clipping techniques. In Ref. [8], signal scrambling technique has been proposed as it reduces PAPR in UFMC. UFMC is adequate for 5G communication, but it also suffers from PAPR, similar to OFDM.

In Ref. [9], the performance of the UFMC system is analysed. Dolph Chebyshev, Hamming and Hanning filtering is used for the generation of UFMC signals. This paper [10] deals with the study of the effect of variation in the number of sub-channels (M) and the overlapping factor (K) of FBMC. In Ref. [11], the authors have proposed multi-carrier modulation such as FBMC and OFDM to address the problem of rate adaptive (RA) for asynchronous cellular networks. The authors [12] have proposed a new key PHY layer technology to increase robustness and reduce latency. The proposed PHY layer technology facilitated an efficient and scalable air interface supporting the 5G drivers. In Ref. [13], inter-carrier interference (ICI) is overcome by using UFMC. In the proposed UFMC, the filtering operation is performed to a group of sub-carriers to reduce the out-of-band side-lobe level and this makes the UFMC scheme to outperform the OFDM. In Ref. [14], from the simulation results, the authors showed that the OOB power of each system increases, and that such increase is the highest in the FBMC system, by evaluating the spectrum characteristics and BER performance of the waveform.

20.3 Proposed UFMC System with Prototype Filter

In this section, a UFMC system with different prototype filters, such as time-limited filters, band-limited filters and localized filters, is proposed. For comparison, an OFDM system is also designed with the same prototype filters.

20.3.1 Prototype Filter

In UFMC, the filtering operation is performed for a group of sub-carriers, which reduces the filter length and complexity. The filter with reduced filter length shapes the waveforms and improves the resistance of inter-carrier interference (ICI) by suppressing the side lobe between the resource block. The following are the three major types of filters that are used for filter bank. Time-limited filters (Rectangular filter, PHYDYAS filter), band-limited filters (Root-Raised-Cosine filter) and localized filters (Hermite filter).

20.3.2 UFMC System Model

Here, in UFMC, filtering is performed for a group of sub-carriers, unlike in FBMC, where filtering is performed for each sub-carrier. In UFMC, the total transmitter bandwidth is split into B number of sub-bands, and each sub-band has k number of sub-carriers. This is shown in Fig. 20.1. The filtered output from each sub-band is added in an adder unit and then it is transmitted into the channel. Eq. (20.1) shows the output of each filter bank.

$$X_K = \sum_{(i=0)}^B F_{(i,k)} V_{i,k} S_{i,k} \tag{20.1}$$

Where $S_{i,k}$ represents symbol-mapped data blocks, $F_{i,k}$ represents the Filter, $V_{i,k}$ represents IFFT to eplitz matrix.

The receiver side of the UFMC is shown in Fig. 20.2. The data signals are processed and are given to the filter bank. After this, the filtered signals are converted back to parallel form for further processing. The signal is de-mapped and the data is reconstructed at the receiver end.

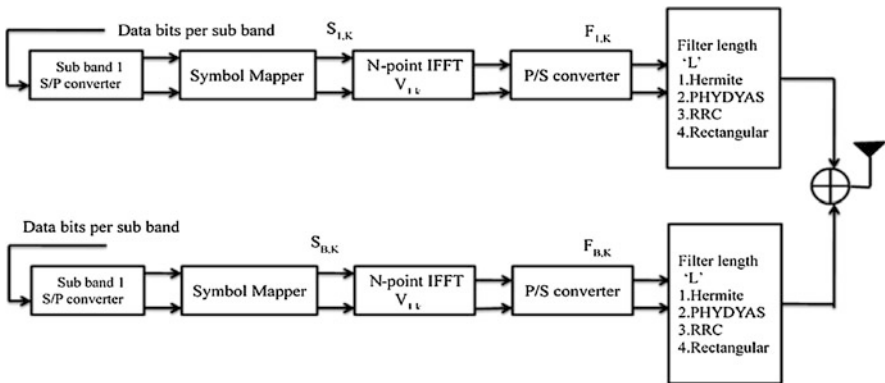


Fig. 20.1 Block diagram of UFMC transmitter

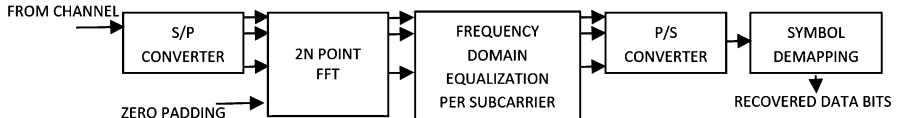


Fig. 20.2 Block diagram of UPMC receiver

20.4 Results and Discussion

In this section, the performance of the UPMC system such as Power Spectral Density (PSD), Block Error Rate (BLER), Bit Error Rate (BER) and Peak-to-Average Power Ratio (PAPR) are analysed and compared with the OFDM system. Both in UPMC and OFDM, prototype filters are considered for performance analysis.

20.4.1 Parameters for Simulation

To compare the proposed UPMC scheme with OFDM technique, a MATLAB platform has been used. The proposed technique is also compared with OFDM using various prototype filters. The simulation parameters are (i) the FFT size is 512, (ii) the number of sub-carriers is 12, (iii) sub-carrier spacing is 15KHZ, (iv) the SNR range is $-5,30$ dB, and (v) the symbol size is 16.

20.4.2 Performance Analysis of the Hermite Filter

20.4.2.1 PSD Evaluation

To illustrate the efficiency of UPMC over OFDM, the PSD of both systems are estimated and simulated using the Hermite filter. The Hermite filter minimizes side-lobe power and the ICI effect, and decreases interference with the neighbouring transmission system. Figure 20.3a shows that the spectral regrowth of UPMC is $(-390$ dB), which is less compared to the spectral regrowth of OFDM, which is $(-110$ dB). It provides 72% improvement compared to OFDM.

20.4.2.2 BLER Analysis

Figure 20.3b shows the BLER value measured for different SNRs for both UPMC and OFDM. The measured values of BLER are 0.000375 for UPMC and 0.001 for OFDM, and it offers a 63% improvement compared with OFDM.

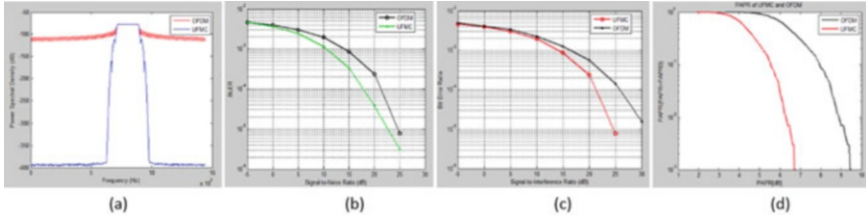


Fig. 20.3 Performance analysis of UFMC and OFDM using Hermite filter. (a) PSD analysis (b) BLER analysis (c) BER analysis (d) PAPR analysis

20.4.2.3 BER Analysis

Figure 20.3c shows the BER for different values of SNR for both UFMC and OFDM. From Fig. 20.3c, it is seen that the BER for UFMC is better than that for OFDM. The BER value has been calculated for the signal-to-noise ratio of 15 dB and for 16-QAM. The BER for UFMC is 0.00025 and that for OFDM is 0.00075, and UFMC offers 67% better performance than OFDM.

20.4.2.4 PAPR Analysis

Figure 20.3d shows that UFMC attains a low Peak-to-Average Power Ratio (PAPR) compared with OFDM. At a CCDF of 10^{-3} , the UFMC system has a PAPR of 6.8 dB, whereas for the OFDM system, this value is 9.6 dB. The UFMC system has 29% better performance compared to OFDM.

20.4.3 Performance Analysis of PHYDYAS Filter

20.4.3.1 PSD Evaluation

To illustrate the efficiency of UFMC over OFDM, the PSD of both systems are estimated and simulated using the PHYDAS filter. From Fig. 20.4a, it is seen that the spectral regrowth of UFMC is very low (-300 dB) compared with that of OFDM (-110 dB). UFMC has 63% better performance than OFDM.

20.4.3.2 BLER Analysis

Figure 20.4b shows that the BLER of UFMC has better performance than that of OFDM for 16QAM modulation and SNR value of 15 dB. The BLER value for

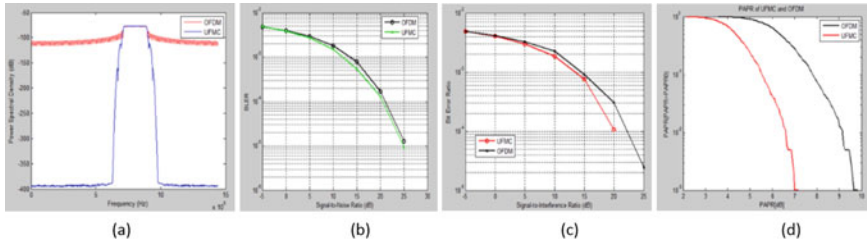


Fig. 20.4 Performance analysis of UFMC and OFDM using Root-Raised-Cosine (RRC) filter (a) PSD analysis (b) BLER analysis (c) BER analysis (d) PAPR analysis

UFMC is 0.000468 and that for OFDM is 0.00075, and UFMC has 93% better performance than OFDM.

20.4.3.3 BER Analysis

The BER analysis for the UFMC and OFDM systems for the PHYDYAS filter is shown in Fig. 20.4c. From Fig. 20.4c, we clearly see that the BER of UFMC is improved compared to the value of OFDM using the PHYDYAS prototype filter. For the signal-to-noise ratio of 15 dB, the measured BER value for UFMC is 0.001 and that for OFDM is 0.002812. UFMC offers 33% enhanced performance than OFDM.

20.4.3.4 PAPR Analysis

Figure 20.4d shows that for a CCDF of 10^{-3} , the UFMC system has a PAPR of 6.7 dB, whereas for the OFDM system, this value is 9.9 dB. The parameter values are FFT size = 512, bits per symbol = 12 and SNR = -5.30 dB. UFMC provides 32% better performance than OFDM.

20.4.4 Performance Analysis of Root-Raised-Cosine (RRC) Filter

20.4.4.1 PSD Evaluation

From Fig. 20.5a, it is seen that the spectral regrowth of UFMC is -390 dB, very minimal compared with that of OFDM, which has a value of -110 dB. UFMC offers 72% better performance than OFDM.

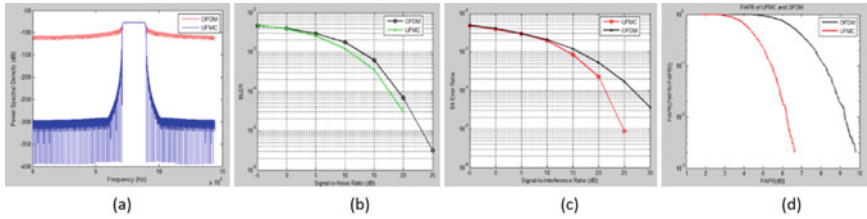


Fig. 20.5 Performance Analysis of UFMC and OFDM using PHYDYAS filter. (a) PSD analysis (b) BLER analysis (c) BER analysis (d) PAPR analysis

20.4.4.2 BLER Analysis

Figure 20.5b shows that the BLER of UFMC is much less than that of OFDM using the RRC filter. For the SNR range value of 15 dB, the measured BLER value for UFMC is 0.000625 and that for OFDM is 0.001. UFMC offers 38% better performance than OFDM.

20.4.4.3 BER Analysis

The BER analysis is shown in Fig. 20.5c. As the order of modulation increases, the BER increases. For the SNR value of 15 dB and QAM modulation of 16, the BER value for UFMC is 0.000875 and that for OFDM is 0.00125. UFMC offers 33% better performance than OFDM.

20.4.4.4 PAPR Analysis

The PAPR analysis for UFMC and OFDM is shown in Fig. 20.5d. As the bits per symbol increases, the PAPR decreases. At a CCDF of 10^{-3} , the UFMC system has a PAPR of 7.2 dB, whereas for the OFDM system, this value is 9.8 dB. UFMC offers 27% better performance than OFDM.

20.4.5 Performance Analysis of Rectangular Filter

20.4.5.1 PSD Evaluation

From Fig. 20.6a, it is seen that OFDM attains better spectral regrowth. The spectral regrowth of OFDM is -120 dB and that for UFMC is -100 dB. Here, UFMC has 10% lower performance than OFDM.

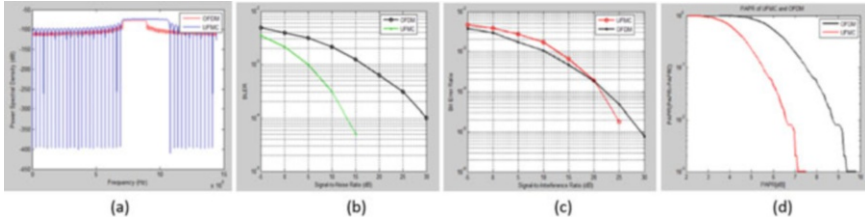


Fig. 20.6 Performance analysis of UFMC and OFDM using Rectangular filter. (a) PSD analysis (b) BLER analysis (c) BER analysis (d) PAPR analysis

20.4.5.2 BLER Analysis

Figure 20.6b shows that the BLER of UFMC is better than that of OFDM using the rectangular filter for a QAM modulation of 16 and an SNR value of 15 dB. The BLER value for UFMC is 0.000625 and that for OFDM is 0.002812. UFMC provides 78% better performance than OFDM.

20.4.5.3 BER Analysis

From Fig. 20.6c, it is seen that the BER of UFMC is better than that of OFDM using the rectangular prototype filter for a signal-to-noise ratio of 15 dB. The BER for UFMC is 0.00075 and that for OFDM is 0.00625. UFMC offers 88% better performance than OFDM.

20.4.5.4 PAPR Analysis

Figure 20.6d shows that for a CCDF of 10^{-3} , the UFMC system has a PAPR of 7.5 dB, whereas for the OFDM system, this value is 9.8 dB. UFMC has 23% better performance than OFDM. The parameter values considered are FFT size = 512, bits per symbol = 12 and SNR range of -5, 30 dB.

20.5 Conclusion

In this chapter, the UFMC system has been designed with new prototype filters such as the Hermite filter, the PHYSIDIAS filter, the RRC filter and the Rectangular filter. The UFMC system has been analysed for the parameters PSD, BLER, BER, PAPR and compared with the OFDM system with the proposed prototype filters. As UFMC uses prototype filters for filtering a group of sub-carriers, its complexity is reduced compared with FBMC. The analysis shows that the Hermite filter provides 72%, 63%, 67% and 29% better performance than OFDM for the parameters PSD, BLER, BER and PAPR, respectively. The PHYDIAS filter provides 63%, 93%, 33% and

32% better performance, respectively, than OFDM for the same parameters. The RRC filter provides 72%, 38%, 33% and 27% better performance, respectively, than OFDM. The Rectangular filter offers a 10% reduction in PSD compared with OFDM. For the other parameters—BLER, BER and PAPR—the Rectangular filter provides 78%, 88% and 23% better performance, respectively, than OFDM.

References

1. An C, Kim B, Ryu HG (2016) Waveform comparison and nonlinearity sensitivities of FBMC, UFMC and W-OFDM systems, NETCOM, NCS, WiMoNe, GRAPH-HOC, SPM, CSEIT – 2016, pp 83–90
2. Kishore KK, Umarand PR, Naveen VJ (May 2017) Comprehensive analysis of UFMC with OFDM and FBMC. *Indian J Sci Technol* 10(17):1–7
3. Farhang-Boroujeny B (2011) OFDM versus Filter Bank Multicarrier. *IEEE Signal Process Mag* 28:92–112
4. Xi Z, Jiay M, Chen L, May J, Qiu J Filtered-OFDM — enabler for flexible waveform in the 5th generation cellular networks. *IEEE Globecom*, San Diego, CA, Dec 2015
5. Van Eeckhaute M, Bourdoux A, De Doncker P, Horlin F (2017) Performance of emerging multi-carrier waveforms for 5G asynchronous communications. *EURASIP J Wireless Commun Netw* 2017(29):2–15
6. Jafri AR, Majid J, Sham MA (2017) Hardware complexity reduction in universal filtered multicarrier transmitter implementation. *IEEE Access* 5:13401–13408
7. Pooja R, Silki B, Himanshu M (2018) Hybrid PAPR reduction scheme for universal filter multi-carrier modulation in next generation wireless systems. *Res Dev Mater Sci* 2(5). RDMS.000549). <https://doi.org/10.31031/RDMS.2018.02.000549>
8. Kumar PP, Kishore KK (2016) BER and PAPR analysis of UFMC for 5G Communications. *Indian J Sci Technol* 9:S1. <https://doi.org/10.17485/ijst/2016/v9iS1/107820>
9. Geng S, Xiong X (2015) UFMC system performance analysis for discrete narrow-band private networks, *IEEE 6th international symposium on microwave, antenna, propagation, and EMC technologies (MAPE)*, <https://doi.org/10.1109/MAPE.2015.7510319>
10. Kang AS, Vig R (2015) Simulation analysis of prototype filter Bank multicarrier cognitive radio under different performance parameters. *Indonesian J Electr Eng Inf (IJEI)* 3(3):157–166
11. Denis J, Pischella M, Le Ruyet D (2014) A generalized convergence criterion to achieve maximum fairness among users in downlink asynchronous networks using OFDM/FBMC. *IEEE Commun Lett* 18(11):2003–2006
12. Wunder G, Jung P, Kasparick M, Wild T, Schaich F (2014) 5GNOW:non-orthogonal, asynchronous waveforms for future Mobile applications. *IEEE Commun Mag* 52(2):97
13. Vakilian V, Wild T Universal-filtered multi-carrier technique for wireless systems beyond LTE, 2013 *IEEE Globecom Workshops*
14. An C, Kim B, Ryu HG (2016) Waveform comparison and nonlinearity sensitivities of FBMC, UFMC and W-OFDM systems, Eighth international conference on networks & communications, December 2016

Chapter 21

Fully Convolved Neural Network-Based Retinal Vessel Segmentation with Entropy Loss Function



V. Sathananthavathi, G. Indumathi, and A. Swetha Ranjani

Abstract The eye is the exclusive organ for the sense of sight in humans. Morphological changes in vascular diameter and branching pattern of retinal vessels lead to blindness. Segmentation of retinal vessels is done to analyse these morphological changes in retinal vessels. However, due to the presence of illumination, multiplex distribution of blood vessels, and low contrast between target and background, the task of segmentation of retinal blood vessels is highly challenging. In this chapter to segment retinal blood vessels, we propose a method based on fully convolutional neural networks and pixel classification with cross-entropy function to avoid the class imbalance problem. Our proposed architecture of fully convolutional neural networks combines the output of each stage to learn the hard samples. The cross-entropy loss function is performed to avoid misclassification of vessels.

Keywords Retinal images · Local normalization · Convolutional neural network · Entropy loss function

Abbreviations

CNN	Convolutional Neural Network
CLAHE	Contrast Limited Adaptive Histogram Equalization
DRIVE	Digital Retinal Images for Vessel Extraction

V. Sathananthavathi (✉) · G. Indumathi · A. Swetha Ranjani
Mepco Schlenk Engineering College, Sivakasi, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_21

21.1 Introduction

Retina is a light-sensitive layer that lines the back of the eye and is affected due to vascular occlusion. Ophthalmologic diseases, such as diabetic retinopathy, or other diseases, such as hypertension and arteriosclerosis, are the main causes of vascular occlusion, which leads to blindness. These ophthalmologic diseases are related to morphological changes in the vascular diameter and branching pattern, leading to blindness. There are many people in the world who suffer from blindness due to retinal diseases that lead to morphological changes in retinal blood vessels. The important requirement for the diagnosis of retinal diseases involves segmentation of retinal blood vessels. Therefore, accurate segmentation of retinal blood vessels is of greater importance for diagnosis and treatment of retinal diseases. Retinal images such as two-dimensional (2-D) and three-dimensional (3-D) colour fundus images are used for the diagnosis of ophthalmologic diseases. Experts or specialists segment the retinal vessels manually from the fundus images. This leads to time-consuming and improper segmentation of vessels. Hence, there is a need for automatic segmentation of retinal vessels for reliable and robust segmentation. Segmentation of retinal blood vessels has already been proposed using various methods with the development of computer-aided systems. Salem et al. [1] proposed an unsupervised method of retinal vessel segmentation by using the RADIUS-based CLustering ALgorithm (RACAL) distance-based principle. Wang et al. [2] proposed a segmentation method based on pixel classification in which CNN is used as a feature extractor and employed random forests as a trainable classifier. Marin et al. [3] proposed a method that involves feature extraction for every pixel such as grey-level features and momentum features, and then these features are applied to neural networks for classification. Conventional supervised method involves feature extraction and classification. Feature extraction is a trivial task, because selecting the best feature is important for the segmentation of retinal blood vessels. Feature selection should be prominent in order to properly segment vessels. Many recent approaches or works use convolution neural network for the segmentation of retinal blood vessels. Convolutional neural network extracts the best feature by performing convolution, batch normalization and pooling operations. In general, retinal vessel segmentation is based on vessel tracking, which has been proposed by Yin et al. [4] or the region-growing technique proposed by Fraz et al. [5] by means of morphological operations, matched filter and the complex continuous wavelet transformation. Liskowski et al. [6] proposed patch-wise segmentation of retinal vessels using CNN, which requires more memory and is a time-consuming process. Orlando et al. [7] proposed a method to reduce luminous and contrast variation in retinal images for better segmentation of retinal vessels using local normalization and proposed a CNN architecture that takes input of arbitrary size and produces output with efficient and inference learning. Zhenkiang et al. [8] used a pre-trained network, Alexnet, for the segmentation of retinal blood vessels. Zhexin et al. [9] considered a green channel as it provides better contrast between vessel and the background, and

developed an architecture of neural network with four convolutions and one fully connected layer. Jose et al. [10] proposed a CNN architecture where class balance problem is not taken into consideration, which leads to false segmentation of non-vascular region as vessels. Kai et al. [11] proposed a method for segmentation of retinal blood vessels using multiscale CNN to get a probability map and then proposed a class balance cross-entropy loss function to improve the performance of segmentation.

21.2 Proposed Method

21.2.1 Overview

In the proposed work, the input image is preprocessed to eliminate illumination and low contrast between the vessels and the background. The image is then enhanced to get detailed information about vessel edges. Features are extracted using multiscale CNN and are classified based on pixel classification. Cross-entropy loss function is performed to reduce misclassification. The block diagram of the proposed work is shown in Fig. 21.1.

21.2.2 Green Channel Extraction

The green channel provides a better contrast between vessels and background. So in order to obtain a better contrast between vessels and background, a green channel of image is taken for robust segmentation of blood vessels. The green channel of image is shown in Fig. 21.2.

21.2.3 Local Normalization

Local normalization tends to normalize the local mean and variance of the image. The local normalized image is shown in Fig. 21.3. This is used to normalize the

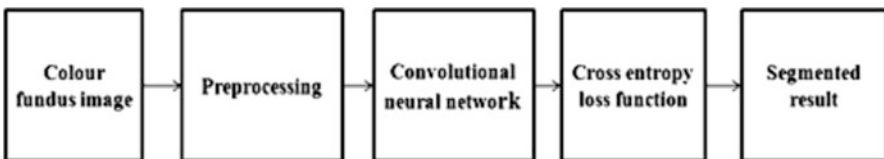


Fig. 21.1 Overview of the proposed work

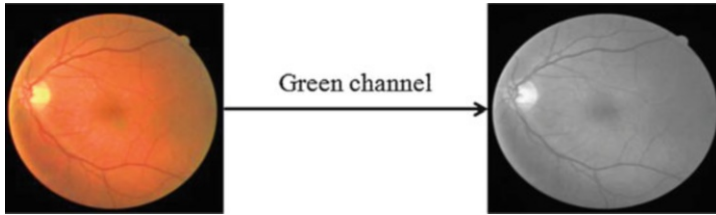


Fig. 21.2 Green channel extraction of input image

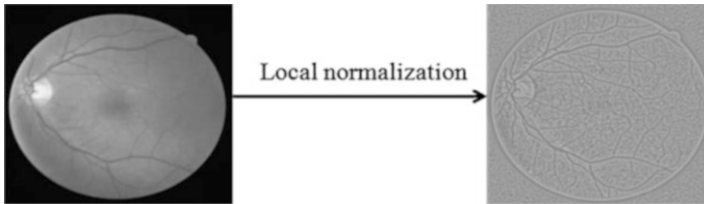


Fig. 21.3 Local normalization of green channel image

illumination of the fundus image so that the vessels can be well distinguished from the background. Local mean and variance are estimated by local spatial smoothing using gaussian filters. Mean of green channel image is subtracted from the green channel image and is divided by variance of green channel image so as to normalize each feature. Smoothing windows, the σ_1 and σ_2 values of gaussian filters, are set as maximum in order to obtain better results. σ_1 and σ_2 values used in the work are four. The gaussian filter used in our work is of size 4×4 . The local normalization is given by Eq. (21.1).

$$f(x, y) = \frac{I(x, y) - m_I(x, y)}{\sigma_I(x, y)} \quad (21.1)$$

$I(x, y)$ is Original image

$m_I(x, y)$ is Local mean of $I(x, y)$

$\sigma_I(x, y)$ is Local variance of $I(x, y)$

$f(x, y)$ is Local normalized image

21.2.4 Contrast Limited Adaptive Histogram Equalization

The green channel image does not have enough contrast variation between vessels and background, which leads to improper segmentation of retinal thin blood vessels. In this proposed work, CLAHE is applied to enhance the contrast of image. CLAHE limits the amplification of histogram by clipping at a predefined value before computing the cumulative distributive function, which is performed by transforming each pixel with a transformation function derived from a neighbourhood. Most

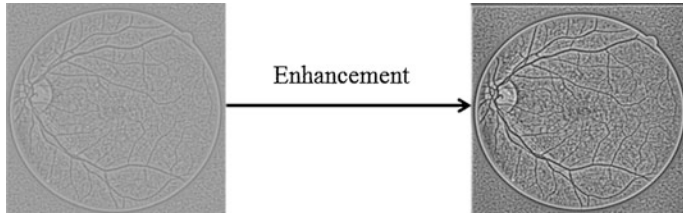


Fig. 21.4 Enhanced image

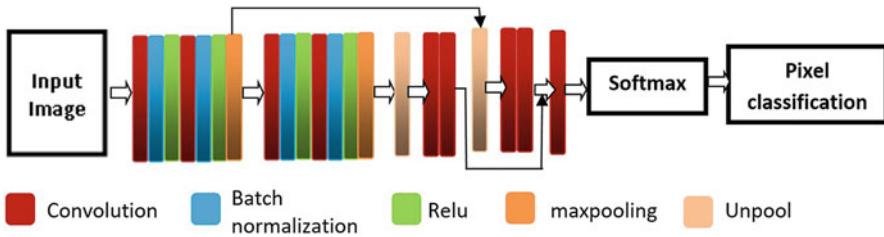


Fig. 21.5 Architecture of convolutional neural network

medical images use this enhancement technique to enable the principal part more visible. The enhanced result of a local normalized image is shown in Fig. 21.4.

21.2.5 Convolutional Neural Network

Convolutional neural network consists of 23 layers. These layers extract features and classify the image. The CNN is composed of an input layer, an output layer and multiple hidden layers. The hidden layer involves a convolution layer, a pooling layer and a batch normalization layer. The proposed architecture for CNN is shown in Fig. 21.5. The first layer is an input layer followed by two stages. Each stage contains hidden layers. The convolution layer in each stage contains filters of size 3×3 with padding on all four sides and striding 1. Batch normalization layer is done between each convolution layer and the ReLU layer to speed up training of convolutional neural networks and reduce sensitivity to network initialization. Sigmoid transfer function diminishes in gradient, whereas rectified linear unit is an activation layer [11], which does not vanish in extremes and so allow effective training of networks with dozens of layers.

In the proposed method 2×2 maxpooling is used. The output of each stage is up-sampled to the size of the input image. The feature map of each stage is merged by convolution. Softmax is performed before pixel classification and the network uses a pixel classification layer to predict the categorical label for every pixel in an input image.

21.2.6 Cross-Entropy Loss Function

The pixel count for vessels and non-vessels is imbalanced for proper classification; only 10% of pixels [11] are labelled for retinal vessels in one fundus image and the majority of pixels are labelled for non-vessels, which results in segmentation of the dominant class. To obtain reliable and robust segmentation, class weighting is done to balance the classes.

The number of classes to classify the image is represented as N , whose corresponding weights are denoted as $w = w^{(1)}, \dots, w^{(N)}$. The total class weight ($T(W)$) is given by Eq. (21.2)

$$T(W) = \sum_{i=1}^N w^{(i)} \quad (21.2)$$

Class frequencies are obtained by dividing each class weight by total class weight. Class frequency F is presented by Eq. (21.3)

$$F^{(i)} = \frac{w^{(i)}}{T(w)} \quad i = 1, \dots, N \quad (21.3)$$

Inverse class weight is obtained by taking the inverse of class frequency. Inverse class weight ($w_I^{(i)}$) is presented by Eq. (21.4).

$$w_I^{(i)} = \frac{1}{F^{(i)}} \quad i = 1, \dots, N \quad (21.4)$$

Pixel classification is done by the final set of layers, softmax and the pixel classification layer. These two layers combine to predict the categorical label for each image pixel. The pixel classification layer is updated with class weight. Better segmentation results are obtained by using class weighting to balance the classes.

21.3 Experiment Evaluation

21.3.1 Dataset

An experiment is conducted on the DRIVE dataset for vessel segmentation of retinal colour fundus images. DRIVE contains 40 images in which 20 images are training images and the remaining 20 are test images. Each training has one ground truth segmented by specialists. Each test image has two ground truths, of which one is a ground truth and other is a gold standard for the ground truth.

21.3.2 Evaluation Metric

Evaluation metrics are used to evaluate the performance of the results. Binary segmentation results contain four cases: true positive (TP), false negative (FN), true negative (TN) and false positive (FP). If the vessel is predicted correctly as a vessel, it is defined as TP and those are wrongly classified as non-vessel pixels are counted as FN. Non-vessel pixels correctly predicted as non-vessel are defined as TN, and the non-vessel pixels wrongly predicted as vessels are defined as FP. Performance is evaluated by four indicators: sensitivity (Se), specificity (Sp) and accuracy (Acc). Se, Sp and Acc are calculated as follows.

$$Se = \frac{TP}{TP + FN} \quad (21.5)$$

$$Sp = \frac{TN}{TN + FP} \quad (21.6)$$

$$Acc = \frac{TP + TN}{TP + TN + FN + FP} \quad (21.7)$$

21.4 Result and Conclusion

21.4.1 Result

Segmentation results of the DRIVE dataset are shown in Fig. 21.6. The evaluation metrics for 20 test images are calculated. Se, Sp and Acc for the DRIVE dataset are tabulated in Table 21.1. The average specificity, sensitivity and accuracy obtained are 95.15%, 94.76% and 95.19%. Proposed method segments most of the vessels.

In our proposed method, the input image is resized to 356×356 . The result obtained at each stage is important for vascular segmentation. Results obtained at the low-level stage is used for detail detection and high-level stage information is used for learning the vascular structure. Therefore, the combination of stages provides better segmentation results. The network is trained for 100 iterations in MATLAB on an Intel i5,8GB system. Epoch initialized for the training phase is 100. The number of epochs is a parameter that defines the number of times the training algorithm has learned the entire training dataset. The training accuracy increased with the number of iterations initially and it remained constant after a certain number of iterations. In Fig. 21.6, the first row shows the result of a low-contrast image, the second row is the result of a highly illuminated image and the third row is the result of a highly illuminated and low-contrast image in the proposed work.

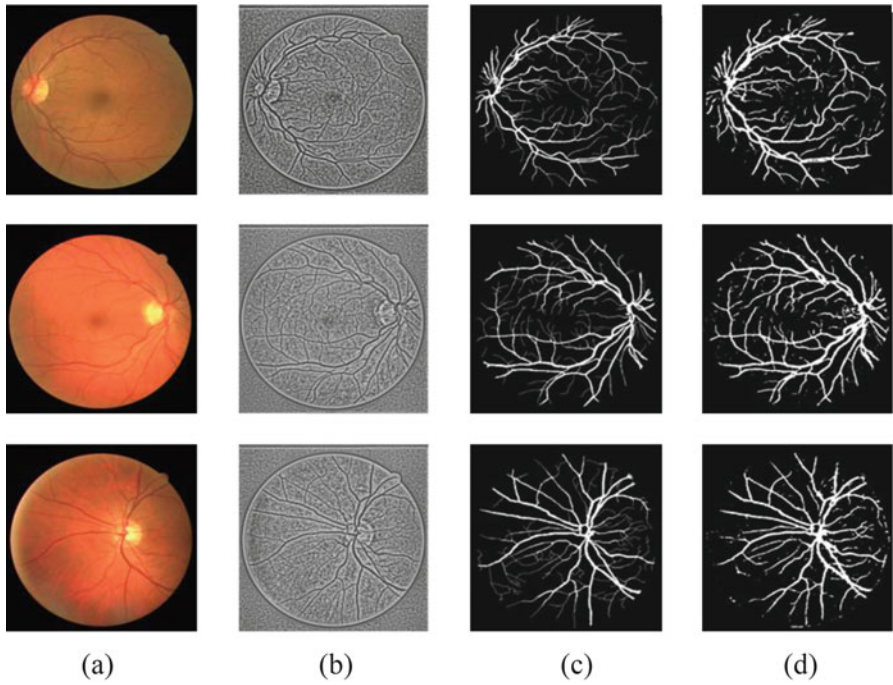


Fig. 21.6 Segmentation result of DRIVE dataset. (a) Colour fundus image. (b) Luminous normalized fundus image. (c) Ground truth. (d) Segmentation result

Table 21.1 Evaluation metrics

Image	Sensitivity	Specificity	Accuracy
1	97.48	93.31	93.58
2	97.46	96.19	96.30
3	94.04	91.82	92.01
4	95.52	95.30	95.32
5	95.97	95.62	95.65
6	96.68	97.12	97.08
7	96.97	95.46	95.57
8	91.01	97.32	96.89
9	94.83	95.68	95.63
10	98.34	94.48	94.71
11	93.17	96.97	96.73
12	96.78	94.90	95.03
13	96.65	94.81	94.95
14	96.18	94.49	94.59
15	98.04	93.43	93.66
16	95.45	95.01	95.04
17	94.11	95.70	95.61
18	92.71	96.45	96.24
19	98.07	94.68	94.88
20	95.73	94.13	94.21
Average	94.76	95.14	95.18

21.5 Conclusion

In the proposed work, the input image is resized, which leads to loss of tiny vessels. Segmentation of retinal vessels using CNN with the inclusion of batch normalization and with improved cross-entropy loss function has optimized the segmentation of retinal blood vessels to some extent. In future, work can be enhanced by exploring efficient loss function.

References

1. Salem SA, Salem NM, Nandi AK (2007) Segmentation of retinal blood vessel using a novel clustering algorithm (RACAL) with a partial supervision strategy. *Med Biol Eng Comput* 45(3):261–273
2. Wang S, Yin Y, Cao G, Wei B, Zheng Y, Yang G (2015) Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 149(PB):708–717
3. Marín D, Aquino A, Gegúndez-Arias ME, Bravo JM (2011) A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants based features. *IEEE Trans Med Imaging* 30(1):146–158
4. Yin Y, Adel M, Bourennane S (2013) Automatic segmentation and measurement of vasculature in retinal fundus images using probabilistic formulation. *Comput Math Methods Med* 2013:260410. 16 p. <https://doi.org/10.1155/2013/260410>
5. Fraz MM et al (2011) Retinal vessel extraction using first-order derivative of gaussian and morphological processing. In: *Advances in visual computing*. Springer, Berlin, pp 410–420
6. Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369–2380
7. Orlando JI, Prokofyeva E, Blaschko MB (2017) A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE Trans Biomed Eng* 64:16
8. Yan Z, Yang X, Kwang-Ting (Tim) Cheng (2018) Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Trans Biomed Eng* 65(9):1912–1923
9. Jiang Z, Zhang H, Wang Y, Ko SB (2018) Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Computerized medical imaging and graphics*. Elsevier.
10. Orlando JI, Prokofyeva E (2018) An ensemble deep learning based approach for red lesion detection in fundus images. *Comput Med Imaging Graph Elsevier* 153:115–127
11. Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369

Chapter 22

Solar Power Forecasting Using Adaptive Curve-Fitting Algorithm



N. Sampathraja, L. Ashok Kumar, R. Saravana Kumar,
and I. Made Wartana

Abstract Electricity is generated from different sources such as thermal, coal, nuclear, solar, wind and so on. The generated electricity is connected to grids for further use. If forecast from renewable energy is available, then the utilization of non-renewable resources could be reduced and so the cost and impact on environment can also be reduced through optimized grid balancing. Solar power is one of the renewable power sources in focus due to the upgradation of photovoltaic technologies and simplified system components. But, the yield out of photovoltaic cells could be strongly influenced by factors such as shadow, cloud, rain, dust, temperature, humidity, panel angle, seasonal effects, panel efficiency and so on. Hence, all these factors need to be considered for forecasting solar power. In this chapter, the ‘Adaptive Curve Fitting’ model is introduced for forecasting solar power, wherein the majority of the algorithm is based on mathematical modelling which considers clear sky reference power, real time power and optionally the weather prediction data.

Keywords Clear sky reference · Time segmentation · Slope of the reference power · Real time power · Cloud effect

N. Sampathraja (✉) · L. Ashok Kumar · R. Saravana Kumar
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: nsr.eee@psgtech.ac.in; lak.eee@psgtech.ac.in

I. Made Wartana
Department of Electrical Engineering, ITN Malang, Malang, Indonesia

Abbreviations

ACF	Adaptive Curve Fitting
NPF	Next Day Forecast Power
SAP	Same Day Actual Power
SFP	Same Day Forecast Power
Tu	Time Unit

22.1 Introduction

Nowadays, electricity is vital for most of the technological advancements and for commercial usage. The major energy consumers are houses, industries and even automobiles, as internal combustion engines are about to be phased out—so the future is electric vehicles. With the increase in demand for electricity, there is huge amount of consumption of non-renewable energy sources such as coal, petroleum and natural gas, which are about to deplete in the near future. Hence, a focus on renewable energy sources such as wind and solar is highly needed at this moment to meet the increasing demand. But predicting the availability of renewable sources such as wind and solar is highly challenging as these sources are highly dynamic in nature. Precise forecasting of these renewable energy sources will enhance the stability and reliability of the grids. Solar power is the ever-trending source for electricity, wherein a lot of research is going on for its efficiency improvement and for forecasting its power [1–3]. With regard to solar power forecasting, there are two basic methods: direct method and indirect method. In the direct method, environmental factors are observed using sensors, satellite images, weather predictions and so on. In the indirect method, probabilistic or time-series mathematical models are derived. In this chapter, a hybrid method called Adaptive Curve Fitting is introduced and its performance is analysed. Full-day clear sky solar panel output is the key consideration for this model, which follows the curve of solar irradiance of a day without any disturbing factors such as cloud, dust and rain, and also this will also indirectly reflect the time-varying performance of the solar panel due to environmental factors such as temperature and humidity. The timely variation of these factors is captured in time-based intervals with the help of the slope of the clear sky solar panel output. The forecasting algorithm further adapts based on the solar power data that is captured in real time and on the weather prediction data, which improve forecasting accuracy by minimizing the impact of random and trend factors. In this work, the full-day duration is split into five segments and the duration of the day is considered between 6 am and 6 pm, which is the typical duration of sunlight availability in India. The algorithm aims to forecast for the same day's remaining hours and for the next 2 days. Forecasting for the next 2 days is based on probabilistic models.

22.2 Proposed Algorithm: Adaptive Curve Fitting (ACF)

It is evident that solar power relies highly on many factors such as solar irradiance, atmospheric temperature, humidity and panel temperature over the day, dust, snow, rain, angle of the panel, cloud, shadow and so on. All these above factors cannot be controlled directly to maintain the desired solar panel output. But the condition of a few factors such as panel angle can be fixed, and the remaining factors can be observed and covered within the mathematical model to forecast solar power. In the proposed model, Adaptive Curve Fitting, the training sets of solar power output are considered to observe the variation of the power output due to the impact of all the above-mentioned factors. It is strongly considered that the solar power output consists of the implicit message about the condition of all the influencing factors, and therefore, the need for direct sensing of these factors is eliminated. On the other hand, the forecasting methodology is based on the clear sky solar power reference [4–6].

The duration of the day is split into five segments and then the slope value is calculated based on the consecutive average reference power for these segments. This slope will be related to the real-time power average of the particular segment and then further power average will be forecast according to the calculated slope data. The list of five segments and their durations are as follows:

- Segment 1: 6.00 am to 8.30 am
- Segment 2: 8.30 am to 10.00 am
- Segment 3: 10.00 am to 2.00 pm
- Segment 4: 2.00 pm to 3.30 pm
- Segment 5: 3.30 pm to 6.00 pm

In each of these segments, the ratio of the impact due to the factors such as cloud impact will vary drastically. Hence, a separate weightage will be assigned based on the reference power output observations. To have control on the random uncertainties and periodic effect on the weather conditions, an option is given in the algorithm to consider the numerical weightage for the real-time weather prediction data from online sources. The weightage will vary from 0 to 1 for weather prediction from full cloud to clear sunny. The next 2 days' forecast is based on the probabilistic model that the next day average power of each segment will be the sum of 70% of the previous day's average power and 30% of the reference average power.

Figure 22.1 explains the Adaptive Curve Fitting algorithm. Realization of this algorithm for solar power forecasting can be performed using a simple hardware setup, as explained in Fig. 22.2.

The hardware setup consists of the following components:

1. 10 W solar panel
2. Voltage and current sensor (MAX471)
3. Real-time clock (DS1307)
4. Arduino Atmega2560 Controller board
5. Wi-Fi module (ESP8266)
6. Mobile phone with 'Blynk' application
7. SD card module
8. 12 V battery

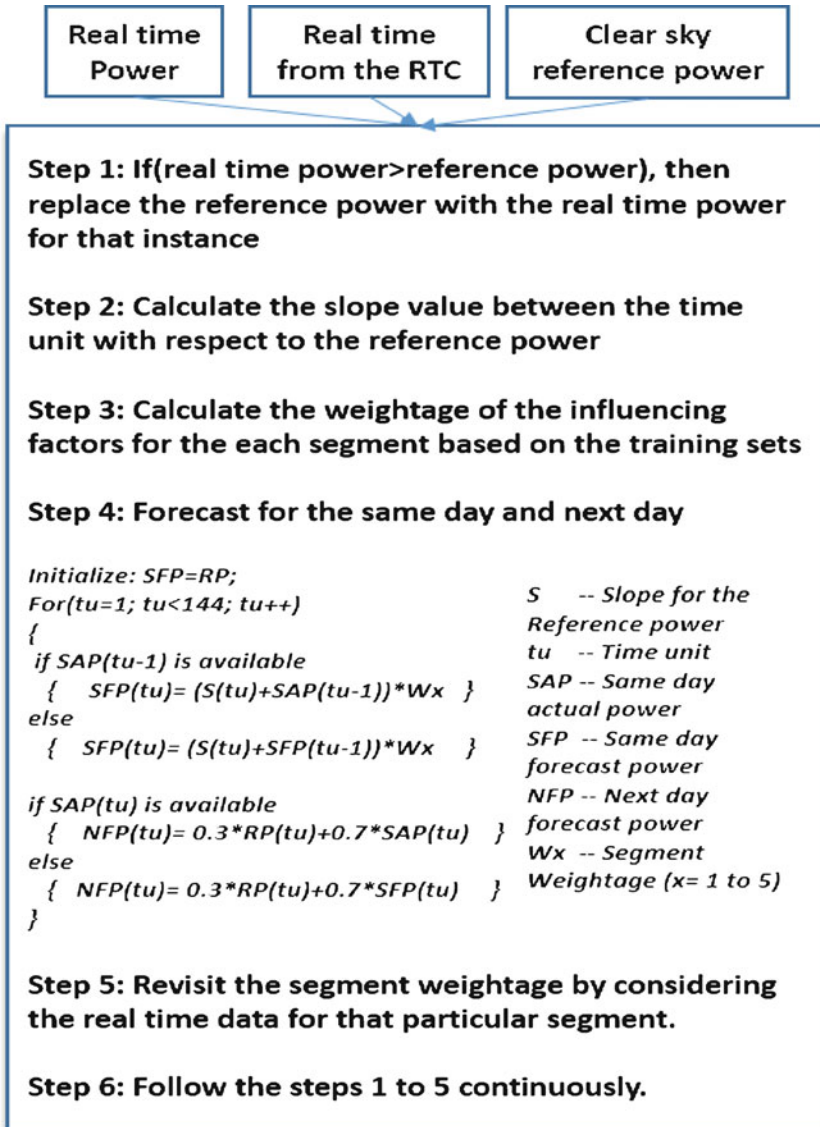


Fig. 22.1 Adaptive curve-fitting algorithm

The panel specification is mentioned in Table 22.1, which is defined for the standard solar irradiance of 1000 W/m^2 . The load resistance is selected at 33 ohms to operate at maximum power point. The panel angle was also fixed, and it is facing south. MAX471 is the voltage and current sensor that is connected to the solar panel

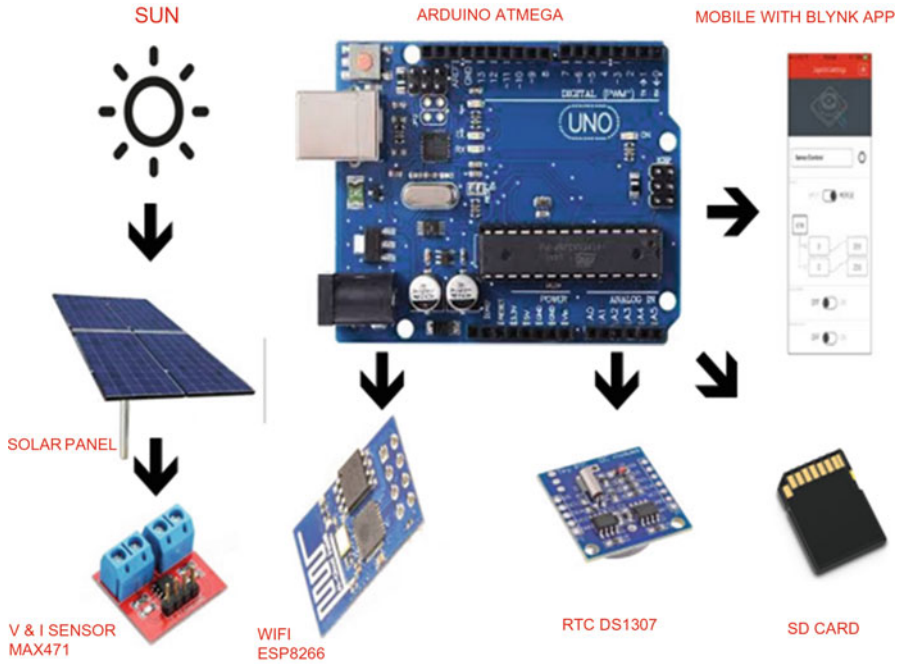


Fig. 22.2 Hardware setup

Table 22.1 Panel specification

Parameter	Value
Material	Polycrystalline
Power (in watts)	10
Open circuit voltage	22
Short circuit current	0.61
Max power point voltage	18
Max power point current	0.56
Fill factor	0.75
Efficiency (%)	6%
No. of cells	36
Weight	1.2 kg
Area	0.162 m2

and the sensor module has an internal sense resistor. It provides analogue signal as the output that indicates the voltage and current flow from the solar panel to the load resistor. DS1307 is the real-time clock module that communicates to the controller board through the I2C protocol. Arduino Atmega2560 is the controller board, which is the open source hardware prototyping module. Two analogue channels, a I2C communication channel, an SPI communication channel and a UART communication channel will be used for this project work. The Wi-Fi module is getting the data

from the controller board through the serial port and it is getting connected to the Wi-Fi hotspot. In the mobile phone, the data can be viewed through the open source Blynk application. The SD card module is used to log the data throughout the day. It is communicating with the controller board through the SPI protocol. A 12 V rechargeable battery is preferred to power the entire hardware system in the long run. The forecasting algorithm software will be loaded into the controller through the Arduino software development environment.

22.3 Result

Capturing of the clear sky reference is a continuous process of recording the maximum power for a particular time instance. Hence, using the few training sets, the reference power is derived based on the moving average method as first-level reference data, and then it would be optimized continuously. Figure 22.3 shows the clear sky reference based on the training sets. The slope value is calculated for each segment. The weightage for the influencing factors is also preset for each segment based on the reference power. The average power values, slope and influencing factor weightage for the reference power is shown in Table 22.2. The slope is nothing but the difference between the average power of the particular segment

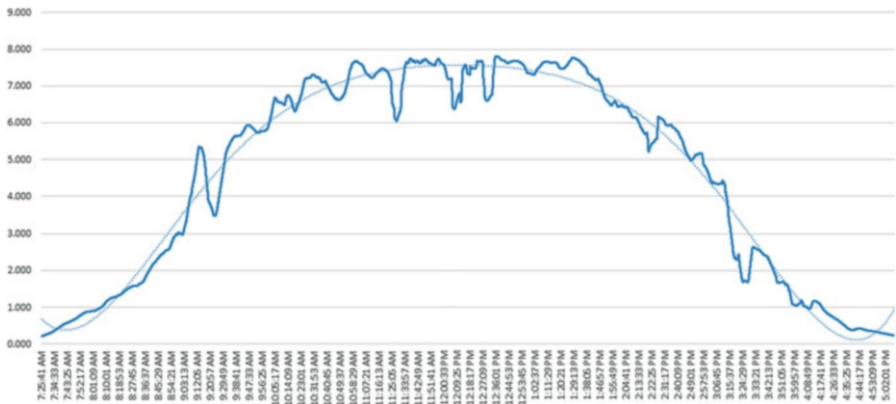


Fig. 22.3 Clear sky reference power with the training set

Table 22.2 Reference power observation

Segment	Reference power (W)	Cloud impact weightage	Slope factor
1	0.88	1	–
2	4.16	0.7	3.28
3	7.23	0.9	3.07
4	4.76	0.55	–2.47
5	0.97	0.8	–3.79

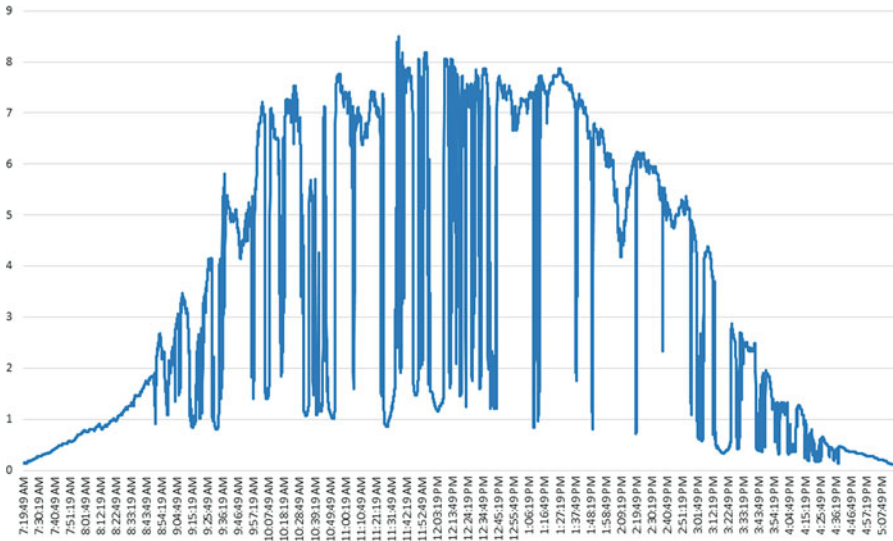


Fig. 22.4 Day 1 Actual power output

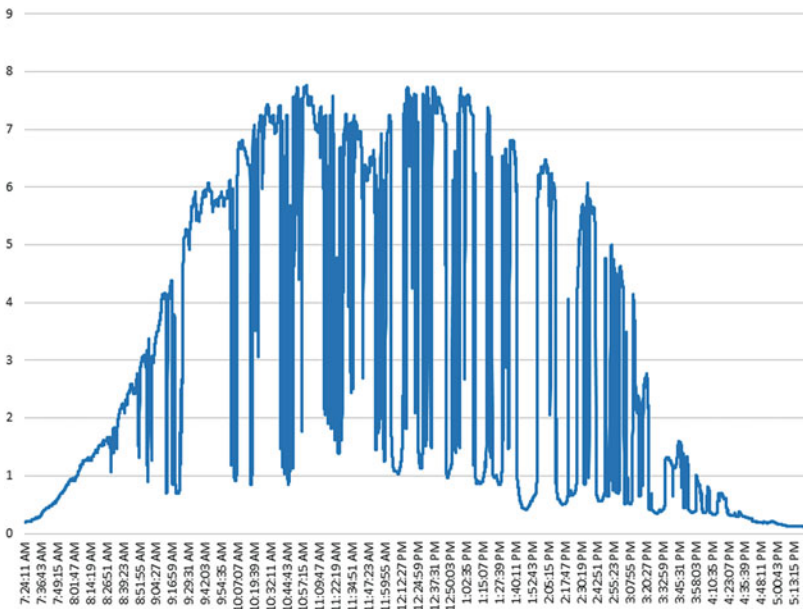


Fig. 22.5 Day 2 Actual power output

and the previous segment. Segments 4 and 5 have negative slope values, which is explicit from Fig. 22.3—that the curve is rolling off. Figures 22.4, 22.5 and 22.6 show the actual data of solar power output for the three consecutive days.

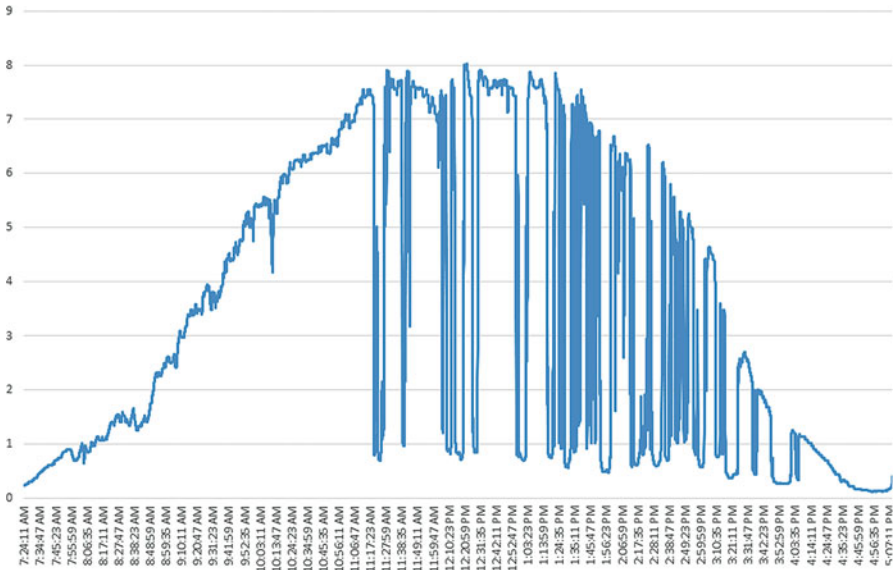


Fig. 22.6 Day 3 Actual power output

Table 22.3 Forecast power vs. actual power for Day 1 at 10 am

Seg.	Day 1 Weather data		Forecast power (W)	Actual power (W)	Deviation (W)
	Factor	Condition			
1	1.00	n.a.	–	0.65	–
2	1.00	n.a.	–	2.28	–
3	1.00	n.a.	5.00	5.61	0.61
4	1.00	n.a.	4.77	4.12	–0.65
5	1.00	n.a.	1.32	0.68	–0.64

Table 22.4 Forecast Power vs Actual Power for Day 2 at 10 am Day 1

Seg.	Day 2 Weather Data		Forecast power (W)	Actual power (W)	Deviation
	Factor	Condition			
1	1.00	n.a	0.72	0.86	0.14
2	1.00	n.a	2.84	3.76	0.92
3	0.80	Partly sunny	4.88	4.71	–0.17
4	0.80	Partly sunny	3.45	2.46	–0.99
5	0.80	Partly sunny	0.61	0.49	–0.12

Tables 22.3, 22.4 and 22.5 show the result of the power forecast data, which is available at 10 am on Day 1. So, the forecast for Day 1 starts from Segment 3 of the actual data of solar power output for the three consecutive days.

Table 22.5 Forecast power vs. actual power for Day 3 at 10 am

Seg.	Day 3		Forecast power (W)	Actual power (W)	Deviation (W)
	Weather data				
	Factor	Condition			
1	1.00	n.a	0.77	0.75	-0.02
2	1.00	n.a	3.24	3.13	-0.11
3	1.00	n.a	5.58	5.62	0.04
4	0.80	Partly sunny	3.07	2.68	-0.39
5	1.00	n.a	0.72	0.78	0.06

With the observation of more data under various climatic conditions, the fine-tuning of the cloud factor is possible, which can improve the accuracy of this model by reducing the deviation between the forecast and the actual power.

The cloud factor calculation, reference power modification and slope calculation are automatic processes with the proposed system. The user can manually feed the weather data through the Blynk application, which will be considered additionally by the algorithm. Else, there will be the usual execution of the algorithm without considering the weather data input. The user can also view the forecasting results in the Blynk mobile application with the help of the proposed hardware system.

22.4 Conclusions

In this chapter, instead of focusing on all the influencing factors of solar panel power output for forecast, an indirect method of time-series analysis has been used with the help of one full-day clear sky data reference and the actual power at a point in time. The Adaptive Curve Fitting will calculate the average power of the day for the different segments by extrapolating the curve by considering the reference curve and other factors such as cloud impact and weather data. The amplitude of the forecast power will depend on the real-time actual power. Thus, the difference between the reference power and the actual power will provide coverage of the influencing factors for the particular time segmentation, which will be extended towards the further segments. The results are easily accessible with the help of Internet of Things technology.

Further improvements are possible with the algorithm by observing more training sets, so that the seasonal climatic effects will also be covered and the cloud factor and slope for the segments could also be optimized with more observations. There could be different reference powers for the different seasons throughout the year. To implement this prototype, necessary adaption is needed with variable panel angle to get the maximum power and with actual load to the solar panel, together with the maximum power point algorithm.

References

1. Cristaldi L, Leone G, Ottoboni R (2017) A hybrid approach for solar radiation and photovoltaic power short-term forecast. *IEEE*, 978-1-5090-3596-0/17/\$31.00 ©2017 IEEE
2. Filipe JM, Bessa RJ, Sumaili J, Tomé R, Sousa JN A hybrid short-term solar power forecasting tool. *IEEE*, 978-1-5090-0191-0/15/\$31.00 ©2015 IEEE
3. Yang C A novel ARX-based multi-scale spatiotemporal solar power forecast model. *IEEE*, 978-1-4673-2308-6/12/\$31.00 ©2012 IEEE
4. Pedro HTC, Coimbra CFM (2012) Assessment of forecasting techniques for solar power production with no exogenous inputs. *Sci Direct* 86:2017–2028
5. Deng F, Su G, Liu C, Wang Z (2010) Global solar radiation modeling using the artificial neural network technique. *Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, March 2010, pp 1–5, 28–31
6. Bacher P, Madsen H, Nielsen HA (2009) Online short-term solar power forecasting. *Sci Dir Sol Energy* 83:1772–1783

Chapter 23

A Review of Electric Vehicle Technologies



P. Ravi Kumar, C. Gowri Shankar, R. Uthirasamy, and V. J. Vijayalakshmi

Abstract Electric vehicles (EVs) have gained remarkable attention due to growing concerns over global warming and the depletion of fossil fuels in the last decade. The propulsion system in EVs comprises electric motors, which are fed by energy storage units through power electronic devices. Due to limitations of conventional energy storage systems in terms of cost, sizing, management, energy and power density, it has become necessary to have an energy generating unit along with the energy storage unit. Effective utilization of energy in EVs can be carried out with the incorporation of advanced power electronics technologies. This chapter reviews the various classifications of EVs, electrical propulsion systems, energy storage systems and energy management systems. This chapter also highlights the various issues to be considered for effective electrification in EVs.

Keywords Electric Vehicle · Energy Management System · Energy Storage System

Abbreviations

EV	Electric Vehicle
ICE	Internal Combustion Engines
ICEV	Internal Combustion Engine Vehicle
HEV	Hybrid Electric Vehicle
AEV	All Electric Vehicles
PEC	Power Electronic Converters
BEV	Battery Electric Vehicle
FCEV	Fuel Cell Electric Vehicle
EPS	Electric Propulsion System

P. Ravi Kumar (✉) · R. Uthirasamy · V. J. Vijayalakshmi
KPR Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

C. Gowri Shankar
KSR College of Engineering, Tiruchengode, Tamil Nadu, India

- ESS Energy Storage System
- PHEV Plugin Electric Vehicle
- UC Ultra capacitor
- EDLC Electric Double-layer Capacitors
- EMS Energy Management System
- RB Rule Based
- RTO Real Time Optimizations

23.1 Introduction

Conventional vehicles utilize gasoline/petroleum products such as petrol or diesel as fuel to propel their internal combustion engines (ICEs), which act as the main source for the energy conversion unit. These vehicles emit greenhouse gases such as carbon dioxide, hydrocarbons, carbon monoxide and nitrogen oxide, which pollute the environment globally. In the last decade, the usage of petroleum products in the transportation sector has raised concerns over the depletion of fossil fuels and the problems related to pollution and climate change.

Reduction in the emission of greenhouse gases is seen as the only major remedy for restoring the natural climatic conditions globally. The transportation sector all over the world has seen a paradigm shift towards electrification of vehicles. Electric vehicles (EVs) are expected to be more reliable, affordable, highly energy efficient and safe with zero or reduced emissions. The prominence of electric vehicles under various dimensions is presented in Fig. 23.1. Electric propulsion systems, which are used in EVs, are considered stiff competition for ICEs with the recent advancements in the field of energy sources, power electronics and electrical-driven technology. This chapter presents an overview of the classification of electric vehicles, energy sources energy management system.

ENVIRONMENT	POLITICS	ECONOMY
<ul style="list-style-type: none"> Climate Change Reduction of CO₂ Emissions Reduction of Noise Emissions 	<ul style="list-style-type: none"> International Specifications for emission limits Low Emissions or Emission free zones 	<ul style="list-style-type: none"> Limited Oil Reserves Rising Prices for Fossil Fuels
TECHNOLOGY	SOCIETY	INFRASTRUCTURE
<ul style="list-style-type: none"> Merits of electric motors compared to internal Combustion engine Increase in Efficiency High Voltage Safety 	<ul style="list-style-type: none"> Growing Mobility Demand for Vehicles with Lower Consumption and Emissions 	<ul style="list-style-type: none"> Comprehensive Infrastructure to supply energy for electric vehicles

Fig. 23.1 Electric vehicle prominence

23.2 Classification of Vehicles

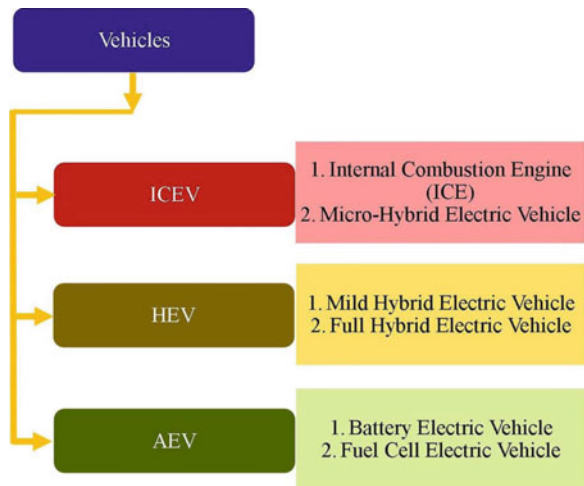
In terms of usage of electrification, the vehicles are categorized into three groups: internal combustion engine vehicles (ICEVs), hybrid electric vehicles (HEVs) and all-electric vehicles (AEVs). Figure 23.2 illustrates the various categories of vehicles.

ICEVs use pure mechanical transmission with petrol or diesel as the main fuel. The use of electrical energy is very minimal, and it is mostly used for starting purpose. HEVs use a combination of electrical and mechanical transmission, and both fossil fuels and electrical energy. Various classifications are available with HEVs, which are discussed below. AEVs are pure electric vehicles that consume electric energy and comprise only the electric drive train and transmission. A comparative overview of the three vehicle categories is shown in Fig. 23.3.

23.2.1 Internal Combustion Engine Vehicles (ICEVs)

These vehicles are propelled by transforming chemical energy to heat energy and kinetic energy by use of a combustion chamber. These vehicles are further classified into conventional ICEVs and micro-hybrid EVs. The major difference is that the former operates without an electric vehicle, whereas the later uses one to start the combustion engine. The operating voltage of the electric motor is around 14 V and power is not more than 5kW (Fig. 23.3).

Fig. 23.2 Classification of vehicles



Vehicle	Energy Source	Propulsion System	Merits	Demerits	Observation
ICEV	Petroleum Products	ICE Based Drive	Reliable Durable Better performance Established technology	Harmful emission Poor fuel economy Less Efficient	Fossil fuel depletion Emission rate is high Fuel economy
HEV	Electrical Energy Petroleum Products	EPS and ICE Based Drive	Very Low Emission Higher Fuel Economy Long Electric Driving Range Durability	Costly and Bulky Increased Component Count Complex Control Algorithm	Size and weight Integration of Component Power Management
AEV	Electrical Energy with Charging Hydrogen enriched fuel	EPS Based Drives	Independent from petroleum Zero Emission Energy Efficient	Higher recharging time Slow transient response Sophisticated electronic controllers	Charging station infrastructure Reliability and cycle life Vehicle performance

Fig. 23.3 Overview of vehicle categories

23.2.2 Hybrid Electric Vehicles (HEV)

These vehicles are propelled by both a combustion engine and an electric motor. Based on the architecture, HEVs are further classified into Mild HEV and Fully HEV. Compared with micro-HEVs, the operating voltage and power of electric motor are 150 V and 7–12 kW in mild-HEVs. However, the rating is high in mild-HEVs, as they can propel the vehicle along with the combustion engine as they use the same shaft.

Fully HEVs have the upper hand in providing better driving performance as they can propel with the use of split power path, that is, they can run either on an electric motor or an ICE or both. This category is further classified into Series Fully HEV, Parallel Fully HEV, Series-Parallel Fully HEV, Complex HEV and Plugin HEV (Fig. 23.4).

Different architectures have different characteristics, which are briefed here. Series Fully HEV comprise of power electronic converters (PECs), which are fed using a battery along with the ICE. With bidirectional operating property of the converter topologies, it is possible to store energy in the battery from the energy generated using the generator. Depending on fuel capacity and generator power, the battery pack can be appropriately designed. It is very much suitable for driving in city conditions. Parallel fully HEVs propel utilizing both electrical power and ICE, along with a mechanical coupler. The vehicles can either run using ICE or EM. These HEVs have better efficiency than series Fully HEVs and can incorporate lower-rated electric motors and batteries as they are very well complimented by ICE. The series-parallel Fully HEVs use two mechanical power couplers as the drive train is powered both electrically and mechanically. These HEVs are costly and complicated even though they incorporate the merits of both parallel and series Fully HEVs. Plug-in hybrid electric vehicles make use of charging infrastructures built to charge the battery from the grid. The charging station and charger configuration are not discussed in the chapter considered. It allows the user to choose the propulsion based on the capacity of the battery installed.

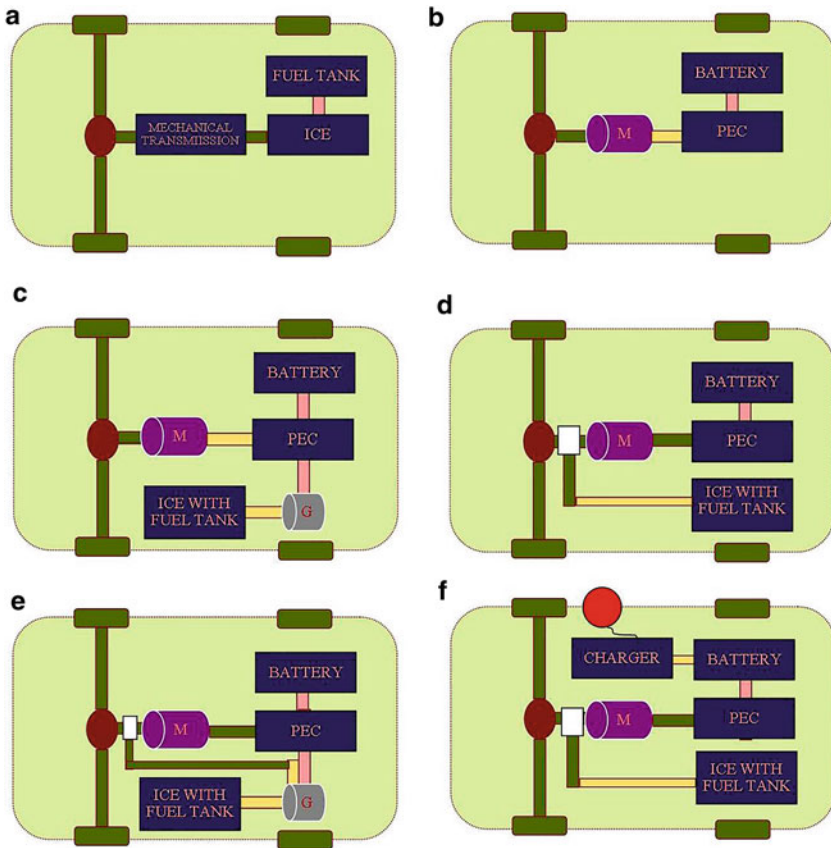


Fig. 23.4 (a) ICEV, (b) BEV, (c) Series HEV, (d) Parallel HEV, (e) Series-Parallel HEV, (f) Plug-in HEV

23.2.3 All-Electric Vehicles

Electric power is the lone source used to propel the vehicle. Battery Electric Vehicles (BEVs) and Fuel Cell Electric Vehicle (FCEVs) are the classifications of AEVs. BEVs use energy-storing devices, mostly batteries, from which energy is converted using a PEC and fed into the motor. Based on the capacity of the battery installed in the vehicle, the driving range of the vehicle can be calculated. Increasing the battery size would increase the weight of the vehicle, thereby reducing its speed. Frequent charging of the vehicles proved to be a demerit compared with HEVs. The FCEVs comprise energy-generating units along with the battery. Hydrogen is the key component, which is being converted using fuel cell technology. Among various energy-generating systems, fuel cell is comparatively better in terms of integration with other energy-storing systems.

23.3 Electric Propulsion System

The core part of any EV is the electric propulsion system (EPS), which comprises components like the energy storage system (ESS), the power electronic converter, the electric motor and the electronic control unit. Based on the architecture of the EV considered, the integration of these components may differ, and it plays a vital role. In order to compete with ICEV, the EPS should take into account issues such as weight of the vehicle, drive train, durability and flexibility, compact packaging and maintenance. Many researches are taking place to design and implement the integration of electric motor and PEC to provide a better driving range. The integration has to focus on issues such as availability of the right motor at affordable cost, promising battery technology and suitable PEC configuration to support efficient transmission.

23.4 Energy Storage System

Energy storage systems consist of both energy sources and energy storing devices, which are attracting the power industries as they help reduce the emission of greenhouse gases. With the advancement of ESS technologies, EV is considered an alternative to ICE, thereby replacing fossil fuels with renewable energy technologies.

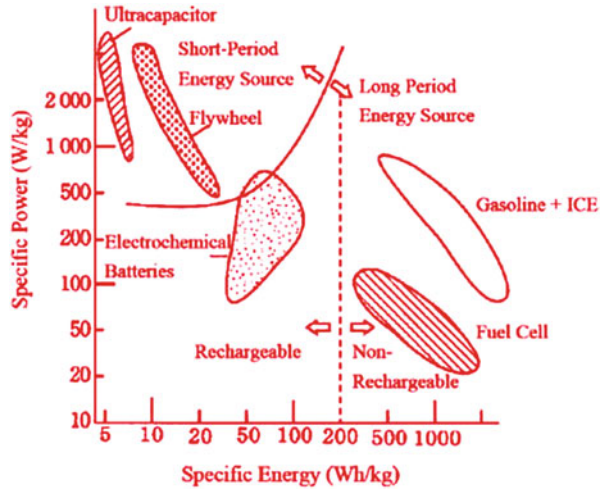
ESS includes energy-generating devices such as fuel cells and other storing devices such as batteries, ultracapacitors and flywheel. ESS can be classified into rechargeable and non-rechargeable technologies.

The operating range of the EV is directly dependent on these technologies and they are illustrated in Fig. 23.5. Comparison of the various technologies is given in Fig. 23.6.

23.4.1 Battery

A battery is a storage device that is made up of electrochemical cells that convert stored chemical energy into electrical energy. Battery capacity and state of charge are the main characteristics of battery, which should be taken into consideration while designing an EV. Other characteristics include high energy density, flexibility, modularity and affordability. Battery requirement is different for different types of EV. BEVs require high-energy-density batteries, HEVs require high-power-density batteries and PHEVs require intermediate battery capacity with energy density like that of a BEV and power density like that of an HEV.

Fig. 23.5 ESS in terms of power and energy



Characteristics	Battery	Ultra Capacitor	Fuel Cell	Flywheel
Mechanism	Chemical	Electrostatic	Chemical	Mechanical
Energy Density	High	Low	Very High	Low
Power Density	Low	Very high	Moderate	High
Charging Time	Hours	Seconds	-	Minutes
Discharging Time	Hours	Seconds	-	Minutes
Life	3-5 years	> 10 years	10–20 years	> 20 years
Efficiency (%)	75-85	85-95	40-60	80-90
Environmental Issues	Disposal	Less	Very Less	Very Less

Fig. 23.6 Comparison of ESS technologies

23.4.2 Ultracapacitor

An ultracapacitor (UC) or super capacitor and a normal capacitor are similar in built, but the former has higher capacitance than a regular capacitor. The ultracapacitor’s characteristics include maintenance-free operation, longer operation cycle life and insensitive to environment temperature variation. The three types of UC technologies established so far are electric double-layer capacitors (EDLC)—carbon/carbon, pseudo-capacitors and hybrid capacitors.

The difference between those UCs is in their energy storage mechanisms and the electrode materials used. In EVTs, ultracapacitors can be used as primary energy devices for power delivery during starting, acceleration and hill climbing, as well as for recovery of braking energy during regenerative braking. It can downsize as well as extend the life of a battery and reduce maintenance and replacement costs.

23.4.3 Flywheel

Flywheel is a storage device that stores and delivers mechanical energy in the form of rotational kinetic energy as well as electrical energy. Limitations of using a flywheel in EVs are that they are heavy weight and costly. Recent advancement in frictionless magnetic bearing, carbon-fibre composite materials, manufacturing technique and sophisticated power electronic controllers has accelerated the development of flywheel energy storage.

23.5 Energy Management

A major challenge in EVs is the development of control and management of power from energy sources to the PEC and from the PEC to various loads. Based on the architecture of the EVs, the energy source may be a single- or multi-input source. Integration of such hybrid sources with the PEC involves understanding of control of PEC components as well as the overall control of the EV for a better driving profile and range. An energy management system (EMS) manages all possible energy resources to feed the power to the ESSs in EVs. An EMS involves low-level component control and high-level supervisory control. This high-level controller usually comprises event-based or time-based conditions that coordinate the component-level operation. There are two classes of control, that is, rule-based (RB) control and optimization approaches control, which are shown in Fig. 23.7.

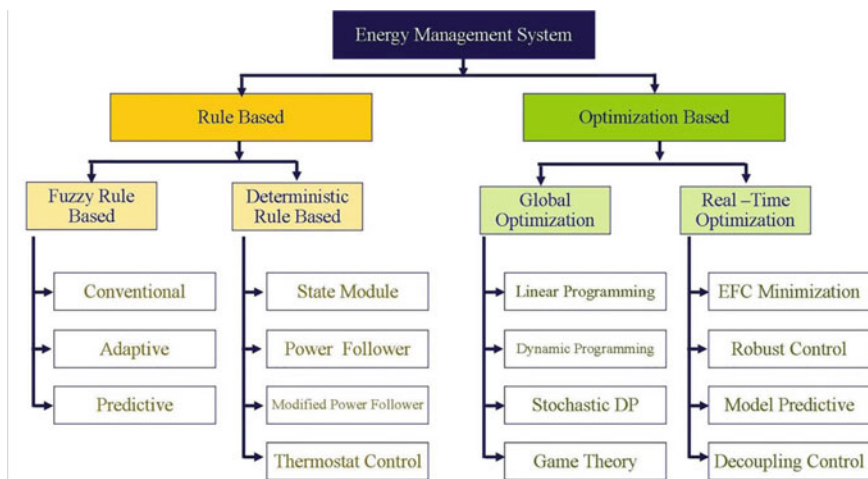


Fig. 23.7 Control and power management strategies of EVs

Rule-based control (RB) is the control system based on human expertise (engineering knowledge), heuristic, intuition, even mathematical model, predefined driving cycles and load-leveiling strategy on vehicles. RB control can be divided into deterministic rule-based methods and fuzzy rule-based methods. Deterministic rule-based use looks up tables (not real-time data) to design deterministic rules. It can be subdivided again into thermostat (on/off) control strategy, power follower (baseline) control strategy, modified power follower (baseline) strategy and state machine-based strategy. However, fuzzy rule-based methods use real-time parameters and suboptimal power split, which are non-linear data and linguistic knowledge to calculate optimal output.

Optimization approach control is based on analytical or numerical operations, which are able, obviously, to minimize the cost function. Optimization approach control can be further classified into global optimization and real-time optimization methods (RTO). Global optimization is based on knowledge of future and past power demands to minimize the cost function through fuel efficiency (fuel consumption) and emissions over a fixed driving cycle. It is useful in design and comparison purposes if implemented together with rule-based strategies. There are a lot of strategies that are categorized under global optimization, for example, linear programming, control theory approach, optimal control, dynamic programming (DP), stochastic DP, genetic algorithm and adaptive fuzzy RB, while real-time optimization consists of equivalent fuel consumption minimization, decoupling control, robust control approach and optimal predictive control. Real-time optimization is based on the system variable at the current data, which is instantaneous cost function. Real-time optimization comprises equivalent fuel consumption minimization strategy (ECMS), decoupling control, robust control approach and optimal predictive control. These control strategies include the drivability of the vehicle.

23.6 Conclusion

This chapter reviews the drive train architecture of EVs with current updates on energy storage as well as energy generation units. It also presents the various electrical drives that can be adopted in EVs. The EMS of an EV is briefed up to introductory level. A comparison with ICEVs is carried out and it can be seen that EVs are the only alternative for reducing emission of greenhouse gases. A few pitfalls seen with AEVs can be seen as opportunities for further research and development.

References

1. Chan CC (2007) The state of the art of electric, hybrid, and fuel cell vehicles. *Proc IEEE* 95 (4):704–718
2. Pesaran A et al (2006) Energy storage requirement analysis for advanced vehicles (fuel cell, mild hybrid, and plug-in hybrid). In: NREL Deliverable Report in Fulfillment of FY2006 August Milestone for Energy Storage Task. Midwest Research Institute (MRI)
3. Mikkelsen KB (2010) Design and evaluation of hybrid energy storage systems for electric powertrains. University of Waterloo, Waterloo
4. Mehrdad Ehsani YG, Ali E (2010) Modern electric, hybrid electric, and fuel cell vehicles, 2nd edn. CRC Press, Boca Raton, p 534
5. Garner IF (1991) Vehicle auxiliary power applications for solar cells. In: Eighth international conference on automotive electronics
6. Emadi A, Williamson SS, Khaligh A (2006) Power electronics intensive solutions for advanced electric, hybrid electric, and fuel cell vehicular power systems. *IEEE Trans Power Electron* 21 (3):567–577
7. Emadi A, Rajashekara K, Williamson SS, Lukic SM (2005) Topological overview of hybrid electric and fuel cell vehicular power system architectures and configurations. *IEEE Trans Veh Technol* 54:763–770
8. Karden E et al (2007) Energy storage devices for future hybrid electric vehicles. *J Power Sources* 168(1):2–11. [78]
9. Chen W, Round S, Duke R (2002) Design of an auxiliary power distribution network for an electric vehicle. In: Proceedings. The first IEEE international workshop on electronic design, test and applications
10. Chau KT, Wong YS, Chan CC (1999) An overview of energy sources for electric vehicles. *Energy Convers Manag* 40:1021–1039
11. Lukic SM, Cao J, Bansal RC, Rodriguez F, Emadi A (2008) Energy storage systems for automotive applications. *IEEE Trans Ind Electron* 55(6):2258–2267
12. Azidin FA, Hannan MA, Mohamed A (2013) Renewable energy technologies and hybrid electric vehicle challenges. *Prz Elektrotech* 89(8):150–156
13. Xing Y, Ma EWM, Tsui KL, Pecht M (2011) Battery management systems in electric and hybrid vehicles. *Energies* 4:1840–1857
14. Rahimi-Eichi H, Ojha U, Baronti F, Chow MY (2013) Battery management system: an overview of its application in the smart grid and electric vehicles. *IEEE Ind Electron Mag*, [June]
15. Emadi A (2011) Transportation 2.0. *IEEE Power Energy Mag* 9(4):18–29
16. Richardson DB (2013) Electric vehicles and the electric grid: a review of modeling approaches impacts, and renewable energy integration. *Renew Sust Energy Rev* 19:247–254

Chapter 24

Gabor Filter-Based Tonsillitis Analysis Using VHDL



P. Nagabushanam, S. Thomas George, D. S. Shylu, and S. Radha

Abstract Image analysis finds application in a wide variety of areas, namely tumour detection, security purpose by monitoring the captured images, diagnosis of early-stage diseases in various parts of the body and so on. Image segmentation plays a major role in image processing to improve the form of an input image for its analysis in further steps. Segmentation is a key factor in image analysis to maintain less computational time and to derive proper meaning in the presence of large distractions and noises in the image. The key challenge in image segmentation is to attain faster computations and low cost without affecting the basic features of the image. This chapter presents several of the segmentation methods used in images. They are (1) Region-Based Segmentation, (2) Threshold-Based Segmentation, (3) Cluster-Based Segmentation and (4) Filter-Based Segmentation. We proposed a new method for image segmentation with Gabor filter bank by orientation of filters in all directions from 0° to 360° . In this chapter, the proposed image segmentation with a Gabor filter is applied for tonsillitis disease-affected image and the simulations using MATLAB and Block Memory Generators (BRAM) using Very High Speed Hardware Description Language (VHDL) in the Xilinx tool are shown.

Keywords Image Segmentation · Gabor filter · Block Memory Generator (BMG) · Tonsillitis · Disease detection · Image to .coe file conversion

P. Nagabushanam
Department of EEE, Karunya Institute of Technology and Sciences, CBE, Coimbatore, Tamil Nadu, India

S. Thomas George · D. S. Shylu · S. Radha (✉)
Department of ECE, Karunya Institute of Technology and Sciences, CBE, Coimbatore, Tamil Nadu, India
e-mail: thomasgeorge@karunya.edu; shylusam@karunya.edu

Abbreviations

RGB	Red Green Blue
BMG	Block Memory Generator
HDL	Hardware Description Language
FPGA	Field Programmable Gate Array
DFU	Diabetes Foot Ulcer
MATLAB	Matrix Laboratory
BRAM	Block Random Access Memory
BROM	Block Read Only Memory
FCN	Fully Convolutional Networks
FCM	Fuzzy c-means clustering
ROI	Region-of-interest
SLIC	Segmentation based lossless image coding

24.1 Introduction

Diabetic foot ulcer (DFU) is a major complication of diabetes, which, if not managed properly, can lead to amputation. DFU can appear anywhere on the foot and can vary in size, colour and contrast depending on various pathologies. The proposed methodology in [1] is a two-tier transfer learning from bigger datasets to train the Fully Convolutional Networks (FCNs) to automatically segment the ulcer and surrounding skin, demonstrating the potential of FCNs in DFU segmentation. A Field Programmable Gate Array implementation of the Gabor-type filter is done by implementing the forward Euler approximation. Simulation and synthesis of results are done using the Xilinx ISE design suite. Verilog HDL is used as a description language for mapping the algorithm in VLSI and hardware implementation is done on SPARTAN-3E FPGA [2].

The authors in [3] carried out a study of segmentation image techniques by using five threshold methods: 1. Mean Method, 2. P-tile Method, 3. Histogram-Dependent Technique (HDT), 4. Edge-Maximization Technique (EMT) and 5. Visual Technique. These techniques were applied on three satellite images to choose base guesses for threshold segmentation image. The aim of this chapter is to develop FPGA realizations of three such algorithms on two FPGA architectures [4]. Case-Based Genetic Algorithm–Location-Dependent Image Classification (CBGA-LDIC) defines multiple location-dependent classifiers, where each classifier is trained by the Gaussian mixture model. CBGA-LDIC decomposes the whole image into some cells, makes a set of cells and then trains classifiers. This method is applied to knee bones [5]. The adaptive integration of the colour and texture helps in the development of complex image descriptors.

Interactive image segmentation with random walker algorithm is applied in arbitrary dimensions on arbitrary graphs [6]. Fuzzy c-means clustering (FCM)

with spatial constraints (FCM_S), which is an effective algorithm suitable for image segmentation. In this chapter [7], a new image segmentation model is used to segment images with severe intensity in homogeneity models using Local Salient Fitting. Segmentation-based lossless image coding (SLIC) was used to get high-resolution image compression [8]. The commonly used functions that are used in CORDIC are implemented on FPGA [9]. The Gabor filter was designed with a new type of multiplier in the ALU called Vedic multiplier and result was obtained much faster [10].

The main challenge in image segmentation for early-stage tumour or disease detection is computational speed. Many segmentation methods have been introduced in order to reduce the number of computations required in image analysis. In this chapter, image segmentation with Gabor filter has been proposed. The novelty lies in the design of filter bank by orientation of filters in all directions for achieving high speed in complex functions computation and low cost. Simulation results are shown for the proposed approach. The rest of this chapter is organized as follows: Sect. 24.2 presents an overview of related segmentation techniques, Sect. 24.3 illustrates the proposed method and Sect. 24.4 provides the simulation results and discussions. Section 24.5 concludes the work done.

24.2 Related Work

24.2.1 *Region-Based Segmentation*

This chapter [5] explains the data bases of sagittal MR images and region-of-interest (ROI) images. The LDIC technology and the genetic algorithm (GA), extended by the case-based technology, are explained. The case-based restart GA for further improvement of classification efficiency is introduced. Segmentation-based lossless image coding (SLIC) is used for all radiographic images. A discontinuity map is obtained during the region-growing procedure. The international bi-level image compression standard Joint Bi-level Image Experts Group (JBIG) was very efficient for coding the discontinuity map and the error image.

In this chapter [6], we can review the relationship of this work (Random Walks for Image Segmentation) to previous approaches. A new approach which is spatially adjacent and deals with distant regions through independent cues is used to construct the image region graph by connecting the graph nodes.

24.2.2 *Threshold-Based Segmentation*

From the international image segmentation method, the specific operation of the process of segmentation method is very diverse and complex, and there is no recognized unified standard. This chapter [11] discusses and compares the four

methods. The first is the threshold segmentation method, followed by the regional growth segmentation, the edge detection segmentation method and the segmentation based on clustering; the last one is the segmentation based on weakly supervised learning in CNN and learns from the shortcomings to analyse better solutions and make future forecasts. The threshold techniques are categorized into two classes: global threshold and local (adaptive) threshold. In the global threshold, a single threshold value is used in the whole image. In the local threshold, a threshold value is assigned to each pixel to determine whether it belongs to the foreground or the background pixel using local information around the pixel. The global threshold has been a popular technique for many years because of its simplicity and easy implementation [3].

In LSF, the image is transformed as the object and the background; a weight factor shows the contrast in the image. The variation degree of local region was also done to extract the distribution information about intensity variation. The salient fitting term is incorporated into the level set method to segment intensity in homogeneous images [7]. A GUI was developed to initialize a fast max-flow-based segmentation algorithm and segment scar with progressive interaction. The hierarchical continuous max-flow (HMF) model was proposed and studied, which is equivalent to the POP model and is implemented using modern convex optimization theory [12].

24.2.3 Cluster-Based Segmentation

The conventional spatial FCM algorithm for image segmentation is introduced and two new low-complexity variants are derived. We obtain a group of kernel fuzzy clustering algorithms with spatial constraints for image segmentation by first replacing the Euclidean norm in the objective functions with kernel-induced non-Euclidean distance measures and then minimizing these new objective functions [13]. In [8], the authors focused on integrate segmentation clustering and atlas construction in a single frame. It is based on atlas-guided segmentation and atlas stratification, which are performed simultaneously; atlas stratification is performed locally, cluster-specific morphological patterns are automatically exposed in a group-wise manner as well as in each individual image, and clinical prior knowledge can be included directly to complement information. This brings the framework in a more data-driven way rather than a hypothesis-driven way [14].

24.2.4 Filter-Based Segmentation

The rotational CORDIC algorithm was implemented using VHDL. The radian values were used as input and the output was converted into decimal form to explain the working of CORDIC [9]. The XILINX Core Generator provides Block Memory Generator one of the IP core to store large images [15]. MATLAB

converts it into .coe files and the Xilinx Core Generator is used to store the coefficient file in a single-port Block ROM by defining the width and depth of the image, which is displayed on a VGA monitor using a Diligent Nexys2 FPGA Board [10]. The automated segmentation of DFU and its surrounding skin by using fully connected networks is proposed in [1]. A two-tier transfer learning method is proposed by training the fully convolutional networks (FCNs) on larger datasets of images and uses it as a pre-trained model for the segmentation of DFU and its surrounding skin. The main contribution to this work [16, 17] demonstrates the global optimization of discrete energy, which can be effectively used for accurate object extraction from N-D snake images (Table 24.1).

The challenge in image segmentation is to obtain a more meaningful form of an image for analysis with low computations. In this chapter, we carried out design and analysis of tonsillitis disease affected image with Gabor filter and CORDIC algorithm for obtaining a .coe file in MATLAB which has pixel values as a look-up table and which can be analysed easily using BRAM in the VHDL Xilinx tool. The novel idea of this work is the orientation of band pass filters in all directions from 0° to 360° to get a Gabor filter bank which is helpful for reducing the number of computations in image segmentation.

24.3 Proposed Image Segmentation Using Gabor Filter and CORDIC Algorithm

In image processing, pre-processing and segmentation should be performed before applying the image for classification. Image segmentation plays a major role in modifying the image for easy analysis in further steps.

The workflow diagram of image segmentation for tonsillitis disease detection is shown in Fig. 24.1.

$$h(x, y) = s(x, y)g(x, y) \quad (24.1)$$

Where $s(x, y)$ is complex sinusoid and $g(x, y)$ is 2D Gaussian-shaped function. Gabor filter is a bandpass filter centred at frequency (x, y) . Radial centre frequency is given by

$$f_c = \sqrt{x^2 + y^2} \quad (24.2)$$

Gabor filters are generated by varying this f_c in all directions from 0° to 360° and all these filters together form a Gabor filter bank.

The Gabor filter can be designed using CORDIC algorithm for faster computation. In CORDIC algorithm, a complex function having real and imaginary parts, namely cosine and sine functions, are converted to kernels. Generally, pixels are stored as 8-bit numbers in a .coe file as a look-up table. CORDIC algorithm reduces the multiplications required and the hardware cost.

Table 24.1 Various models/techniques in segmentation

Ref paper no:	Segmentation/classification method used	Type of filter or model	Applied to images/signals	Applied on (part of the body)	Performance metrics/achievements	Software used
[5, 6]	Region growing	GA (Genetic algorithm), random walks	Knee bone images	Knee bones, radiographic images, colour-texture analysis	Classification efficiency improved, Eigen vector histogram	MATLAB
[3, 11]	Threshold	Weakly supervised learning	Images	Image recognition	To decide foreground or the background pixel	MATLAB
[7, 12]	Level set method, max-flow-based	HMF model	Images	In homogeneous images	Convex optimization theory	MATLAB
[18]	Weighted SVM	Transfer learning	Images	Image recognition		MATLAB VHDL
[8, 13, 14]	Clustering	FCM algorithm, atlas-guided	Brain images	Framework in more data-driven way rather than hypothesis	Minimizing objective functions	MATLAB
[2, 9, 10, 15, 19]	CORDIC algorithm	Urdhva Triyagbhyam – Vedic method, Gabor filter	Images	Block memory generator one of the IP core	Multiplication speed improved	VHDL
[1, 16, 17]	Automated segmentation of DFU	Transfer learning method	Skin images	Skin, extraction from N-D images	Overcome the deficiency of DFU dataset, global optimization of discrete energy	MATLAB
[4]	Integrate segmentation		Snake videos processing	Snakes	Little human interaction	MATLAB

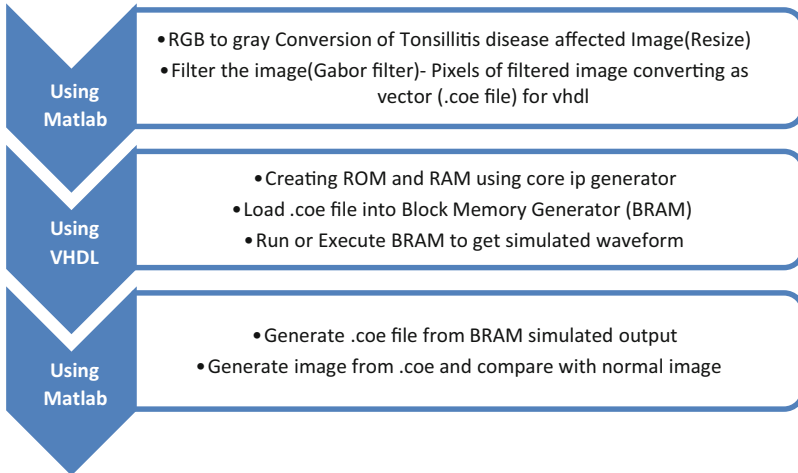
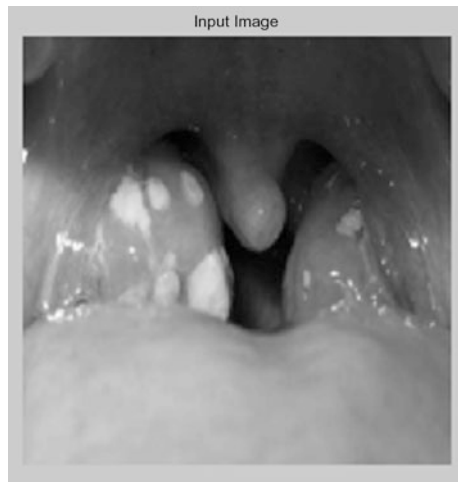


Fig. 24.1 Workflow of proposed image segmentation

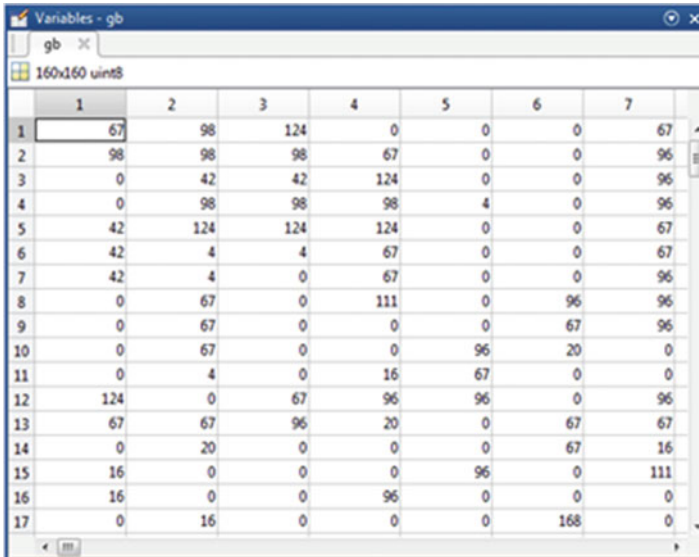
Fig. 24.2 RGB to grey conversion tonsillitis disease-affected image



24.4 Simulated Results and Discussions

The tonsillitis disease-affected image was used as input image for simulation using MATLAB. The simulated results are described in Figs. 24.2, 24.3, 24.4, 24.5, 24.6 and 24.7, in which Figs. 24.2, 24.3, 24.4 and 24.5 are outputs of 3D to 2D conversion, filtered image, .coe file with pixel values and vector form of .coe file for VHDL input, respectively.

Fig. 24.3 Filter the image
(Gabor filter orientation 45°)



	1	2	3	4	5	6	7
1	67	98	124	0	0	0	67
2	98	98	98	67	0	0	96
3	0	42	42	124	0	0	96
4	0	98	98	98	4	0	96
5	42	124	124	124	0	0	67
6	42	4	4	67	0	0	67
7	42	4	0	67	0	0	96
8	0	67	0	111	0	96	96
9	0	67	0	0	0	67	96
10	0	67	0	0	96	20	0
11	0	4	0	16	67	0	0
12	124	0	67	96	96	0	96
13	67	67	96	20	0	67	67
14	0	20	0	0	0	67	16
15	16	0	0	0	96	0	111
16	16	0	0	96	0	0	0
17	0	16	0	0	0	168	0

Fig. 24.4 Pixels of filtered image (.coe file)

24.4.1 Results Obtained in MATLAB

The 2D image obtained after a tonsillitis disease-affected input image applied for RGB to grey level conversion is shown in Fig. 24.2. It is then applied for filtering and we used an orientation of 45° to get a Gabor filter operation. It also uses CORDIC algorithm to reduce the computation time. The filtered image thus obtained is shown in Fig. 24.3.

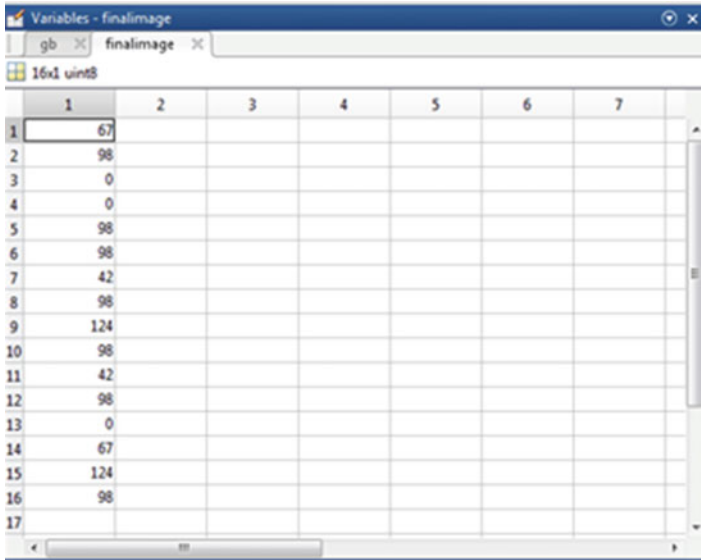


Fig. 24.5 Pixels of filtered image (in MATLAB) converting as vector (.coe file) for VHDL

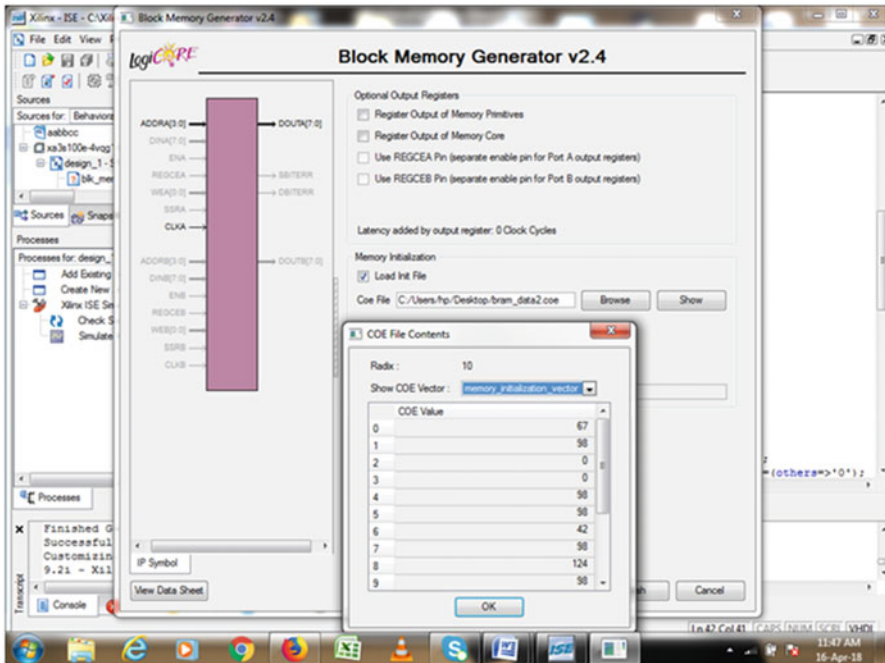


Fig. 24.6 Creating ROM and RAM using core IP generator in VHDL, load .coe file into Block Memory Generator (BRAM)

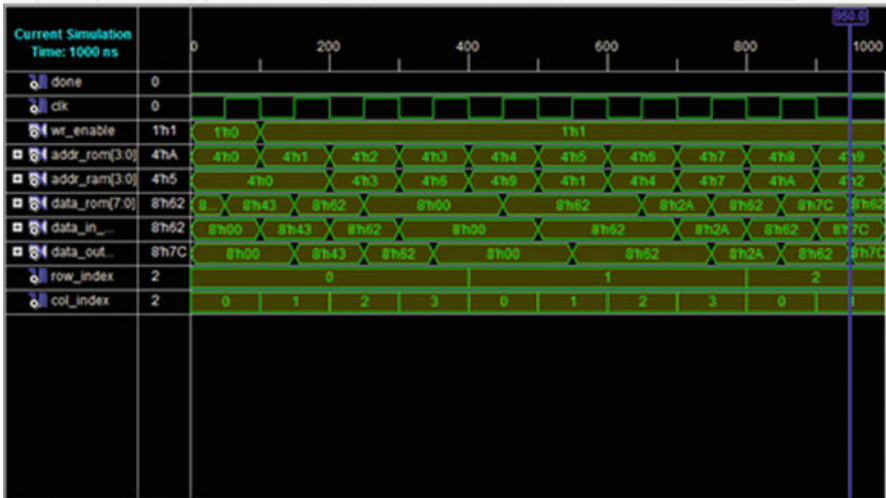


Fig. 24.7 Run or execute BRAM to get simulated waveform

The Excel file shown in Fig. 24.3 with pixel values of filtered image is collected and converted to a pixel vector as shown in Fig. 24.4. It is then given as input for VHDL operations.

24.4.2 Results Obtained Using VHDL

The diagram of core IP generators of ROM and RAM in VHDL is shown in Fig. 24.6. The pixel vector as .coe file from MATLAB is loaded through this ROM and RAM. Then the core generators (BRAM) are executed to get the simulated waveform shown in Fig. 24.7.

24.5 Conclusion

In this chapter, we proposed an image segmentation technique for tonsillitis disease detection to improve the computation speed. The associated theory and design steps were discussed. We showed the orientation of the Gabor filter in all directions (45° orientation shown in this work) to get a filter bank and with CORDIC algorithm, complex functions are computed with less computational time and also remove the limitations of higher number of multiplications required. The .coe pixel file is loaded into IP core generators BRAM (Block Memory Generator) and the simulated results using VHDL are shown.

References

1. Goyal M, Reeves ND, Rajbhandari S, Spragg J, Yap MH (2017) Fully convolutional networks for diabetic foot ulcer segmentation. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff Center, Banff, Canada, October 5–8, 2017
2. Sunitha MK, Harsha BK (2014) Design and implementation of Gabor type filters on FPGA. The International Journal of Engineering And Science (IJES) 3(6):25–31
3. Al-amri SS, Kalyankar NV, Khamitkar SD (2010) Image segmentation by using threshold techniques. J Comput 2(5). ISSN 2151-9617
4. Nelson AE (2000) Implementation of image processing algorithm on FPGA hardware, Nashville, TN
5. Kubota Y, Tsuruta S, Sakurai Y, Kobashi S, Knauf R (2017) Evaluation of a classification method for MR image segmentation. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) Banff Center, Banff, Canada, October 5–8
6. Grady L (2006) Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11), November 2006
7. Min H, Lu J, Jia W, Zhao Y, Luo Y (2018) An effective local regional model based on salient fitting for image segmentation. Neurocomputing. <https://doi.org/10.1016/j.neucom.2018.05.070>
8. Shen L, Member, IEEE, Rangayyan RM (1997) A segmentation-based lossless image coding method for high-resolution medical image compression. IEEE Trans Med Imag 16(3):301–307
9. Ranjan S, Patnaik SK (2015) Design and implementation of CORDIC algorithm using VHDL. Int J Emerg Trend Electric Electron (IJETEE – ISSN: 2320-9569) 11(5):7–10
10. Muralikrishna B, Gnana Deepika K, Raghu Kanth B, Swaroop Vemana VG (2012) Image processing using IP core generator through FPGA. Int J Comp Appl (0975 – 8887) 46(23):48–52
11. Yuheng S, Hao Y. Image segmentation algorithms overview
12. Radha S, Hari Krishna RB, Pandi NP, Varghese S, Nagabushanam P (2018). [Floor planning of 16 bit counter design for health care applications using 180nm technology in cadence tool](#), 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2018
13. Chen S, Zhang D (2004) Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. IEEE transactions on systems, man, and cybernetics—part B: Cybernetics, 34(4), August
14. Ribbens A, Hermans J, Maes F, Vandermeulen D, Suetens P Unsupervised segmentation, clustering and groupwise registration of heterogeneous populations of brain MR images. IEEE Trans Med Imag
15. Nagabushanam P, Radha S, Selvadass S, Joseph KK (2018). [Gabor filter based Image segmentation for disease detection using VHDL](#). 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE, 2018
16. Boykov Y, Funka-Lea G (2006) Graph cuts and efficient N-D image segmentation. Int J Comp Vis 70(2):109–131. Springer
17. Radha S, Mathew J (2017). [Linearization of low noise amplifier for wireless sensor networks](#). Inventive Systems and Control (ICISC), 2017 International Conference, IEEE, 2017
18. Frizhandi AK, Asemanni D (2015) Comparison of images recognition using VHDL and multiclass SVM. Int J Sci Eng Technol Res (IJSETR) 4(9)
19. Rahim N, Islam S, Rokon IR (2015) Design of a modified gabor filter with vedic multipliers using verilog HDL. Int J Inform Electron Eng 5(5)

Chapter 25

Incorporation of Modified Second-Order Adaptive Filter in MFGCI for Harmonic Mitigation of Microgrid



P. C. Keerthiga , G. Gabriel Santhosh Kumar , and S. Hemila Haland

Abstract Power quality is important in the distributed grid system to supply clean and stable power. A multifunctional grid connected inverter (MFGCI) is intended for effective utilization of Distributed Energy Resources and also to provide continuous supply still in deprived power quality situations. The shunt-series switched (SSS) configuration is used for both current and voltage compensation. Therefore, MFGCI is capable of providing series and parallel tie with the grid and load using bidirectional switches. Conversely, the functionality of SSS-MFGCI is imperfect during voltage compensation. In addition, power quality troubles regarded are partial to harmonics. Thus, a Modified Second-Order Adaptive Notch Filter (SOAF) is implemented for multipurpose control of MFGCI and power quality improvement. The control scheme is multipurpose as SOAF is used for grid synchronization, reference signal generation and total harmonic distortion (THD) reduction. The filter technique is adaptive to circumvent the loss occurred. A modified SOAF is used for its improved disturbance rejection capabilities.

Keywords Grid connected inverter · Harmonic distortion · Multifunctional converter · Multipurpose control · Second-order adaptive filter · Power quality · Voltage regulation

Abbreviations

AC	Alternative Current
ANF	Adaptive Notch Filter
DC	Direct Current
DER	Distributive Energy Resources
DG	Distributed Generation
DVR	Dynamic Voltage Restorer

P. C. Keerthiga (✉) · G. Gabriel Santhosh Kumar · S. Hemila Haland
KRCE, Trichy, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_25

GCI	Grid Connected Inverter
MFGCI	Multifunctional Grid Connected Inverter
MG	Microgrid
PCC	Point of Common Coupling
PLL	Phase-Locked Loop
PQ	Power Quality
RES	Renewable Energy Source
SOAF	Second-Order Adaptive Notch Filter
SSS	Shunt Series Switched
STATCOM	Static Synchronous Compensator
THD	Total Harmonic Distortion
UPS	Uninterrupted Power Supply

25.1 Introduction

Modern environmental policies plus requirements of energy have paved way to massive employment of renewable source utilization. Major problems that arise due to the integration of renewable energy sources (RES) to grids are voltage and frequency fluctuations and harmonics [1]. These revisions primarily seek on microgrid control design and grid synchronization methods. The chief aim of such learning is to aid the microgrid as well as distributed supplies to travel in the course of the miniature disturbances in grid attached also to islanded modes. As a result, the usage of a forceful plus unfailing controller and synchronizer is a necessity in microgrids [2]. Different control techniques have been introduced for efficient operation of grid-connected inverters [2–4]. Variable compensation control of power converters only improves current-based PQ [5]. The capability of active power filters depends on grid inductance, along with the position of the compensator [6]. This paved way for multifunctional inverters that provide an effective interface of distributive energy resources and microgrids. They also acts as power quality conditioner and prove to be cost effective [7]. The new control strategies leave information break over compensation as well as control strategies, also in-depth investigation on obtainable power is required [8]. The multi-objective control with respect to power quality service results in issues of limited capacity [9]. Modelling of second-order generalized integrator-based PLL technique for converters shows good performance but shows limited capability in rejecting grid disturbance and oscillatory errors in the presence of DC offset and sub-harmonics [10]. The multipurpose control of grid connected inverters is achieved by using adaptive notch filters (ANF). It acts as a frequency adaptive sequence components extractor. Also, it performs multiple tasks and avoids the usage of PLL [11]. It is most suitable for grid synchronization, acts as a barrier to harmonics and other disturbances in grid signal also offers proficient detection of various components of grid signal [12]. Amongst

various available synchronization procedures together with adaptive techniques, the modified second order adaptive filter (SOAF) is chosen for SSS-configured MFGCI that possess high disturbance rejection capability. Modified SOAF removes the essential components present in deformed grid signal during grid disturbances [13]. The proposal is given in favour of voltage plus frequency fixation in both grid conditions furthermore to make sure of the power production adequacy. The quick and perfect synchronization procedure is chosen along with rapid reacting control structure. Because of the various operations and advantages of an adaptive filter, it has many applications [14–19]. The modified SOAF-supported microgrid system is designed using SIMULINK/MATLAB. By the results of MATLAB simulation, the proposed system is proved to be with high precision. In addition, we can observe balanced three phase output voltage with constant frequency.

25.2 SSS Configured MFGCI

A shunt series switched (SSS)-configured MFGCI is shown in Fig. 25.1. This novel SSS configuration is found to be effective in comparison with all other conventional grid-connected inverters, as it is used to compensate for various quality problems in any power grid by its different modes of operation. Bidirectional switches are used for SSS configuration and are responsible for various operating modes and compensation Schemes [8]. All the four operating modes and their respective functions of the SSS-configured MFGCI are given in Table 25.1. Mode A presents the shunt

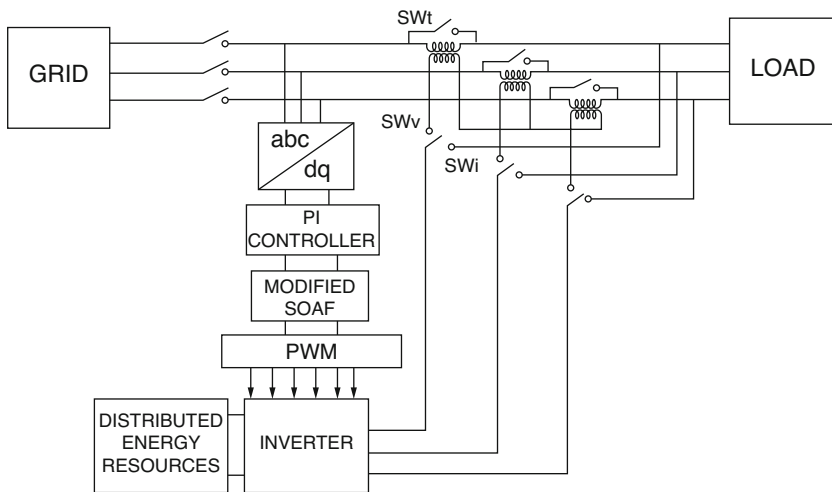


Fig. 25.1 The proposed system with shunt-series-switched MFGCI (SSS-MFGCI) with modified SOAF structure

Table 25.1 SSS-MFGCI modes and functions

Mode	Configured	Switches			Functions
		SW_i	SW_i	SW_v	
A	Shunt	On	On	Off	GCI
B	Shunt	On	On	Off	STATCOM
C	Series	Off	Off	On	DVR
D	Shunt	Off	On	Off	UPS

configuration that provides AC output to the grid from distributed sources DC power. During power quality problems that are current based, Mode B acts as an active power filter, whereas during problems that are voltage based, it acts as STATCOM. Mode C mitigates severe power quality problems that are only voltage based, thus it acts as a dynamic voltage restorer (DVR). Mode D is the uninterruptible power supply (UPS) operation that provides demand power even after the disconnection of the main grid from the load [3].

25.3 Modified Second-Order Adaptive Filter

The modified second-order adaptive filter (SOAF)-based synchronization method is implemented for extraction of harmonic components from the grid signals in grid-connected distributed energy systems. SOAF also aids in the decrement of THD, thus providing power quality enhancement, which is affected by the presence of non-linear loads. The feedback signal for the proposed modified SOAF is the required frequency that is generated from the fundamental component extracted. The analysis shows high disturbance rejection capability in the system and very low settling time followed by satisfactory transient performance and hence found to be effective for fundamental component extraction in comparison with other grid synchronization techniques. The SOAF is modified so as to advance the computational time and for further reduction of total harmonic reduction, and the focus of the modification is to make it suitable for the application of DERs in power grid along with power quality improvement, whose transfer function was modified as,

$$k = g3\hat{\omega}ev$$

The transfer function of input to output is an adaptive band pass filter and the transfer function of input to error is a notch filter.

$$ABPF1(S) = AF1(S)/1 + AF1(S); ANF1(S) = 1 - ABPF1(S)$$

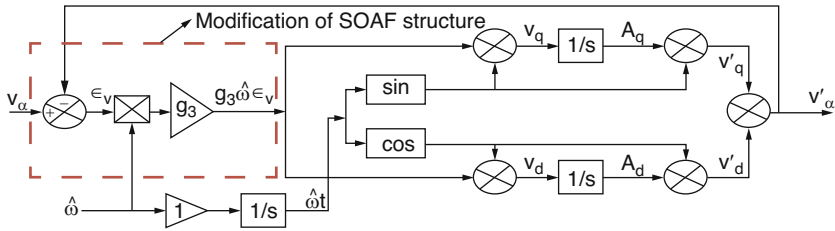


Fig. 25.2 Block diagram of modified second-order adaptive filter

The resolving period of the proposed second-order system is calculated as $T_{s1} = 4.6\tau_1$; here τ_1 is the time constant.

The modified-SOAF capability of filtering is found to be more than satisfactory; however, it takes a relatively high settling period and shows faster reaction for both high and small gain values. Subsequently, bandwidth is solitarily based on gain (g_3) and hence is suitable for the implementation in variable-frequency power systems (Fig. 25.2).

25.4 Simulation Results

The control scheme with suggested SSS-MFGCI is validated with MATLAB Simulink to check the performance. Voltage-quality issues on grid voltage are justified here. In Mode B, the compensating current is injected. Voltage regulation is done by negative sequence current of SSS-MFGCI. However, complete compensation at PCC is not achieved and grid current is unbalanced. That is, the real and reactive power of SSS-MFGCI is oscillating because of the negative sequence current. For this, the reference voltage must be aligned to achieve zero-averaged active power. Thus, the required positive and negative sequence voltage is injected during Mode C. As a result, with less reactive power consumed, both grid and load voltage are sinusoidal and balanced. So, the modified SOAF is the perfect choice for effective operation and for achieving low THD. The simulation results for such a system with SSS-MFGCI and SOAF are shown below. The output shows the intentional voltage disturbance between time period of 0–1 and 1.5–2 at grid voltage even for which the output load voltage is undisturbed, sinusoidal and balanced, and thus the THD at load side reduces (Figs. 25.3, 25.4, 25.5 and 25.6).

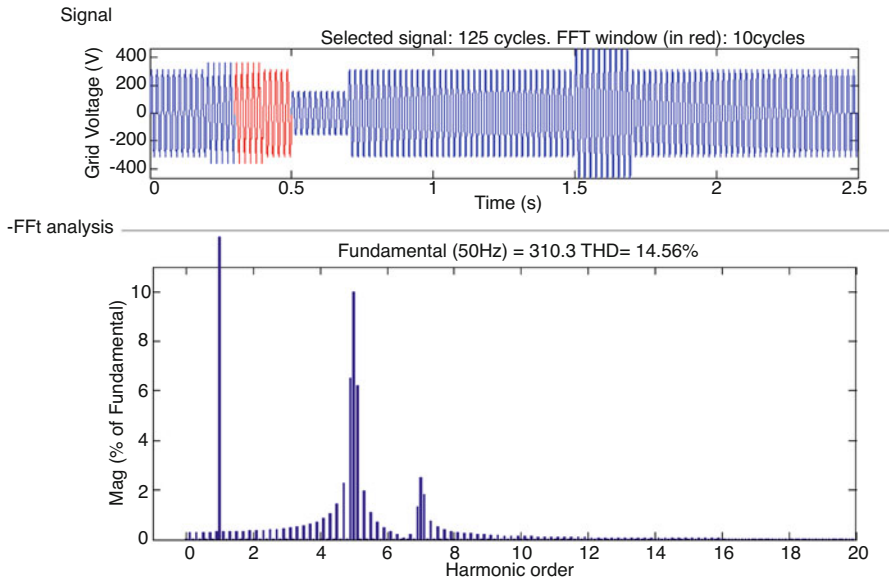


Fig. 25.5 Existing system harmonic distortion

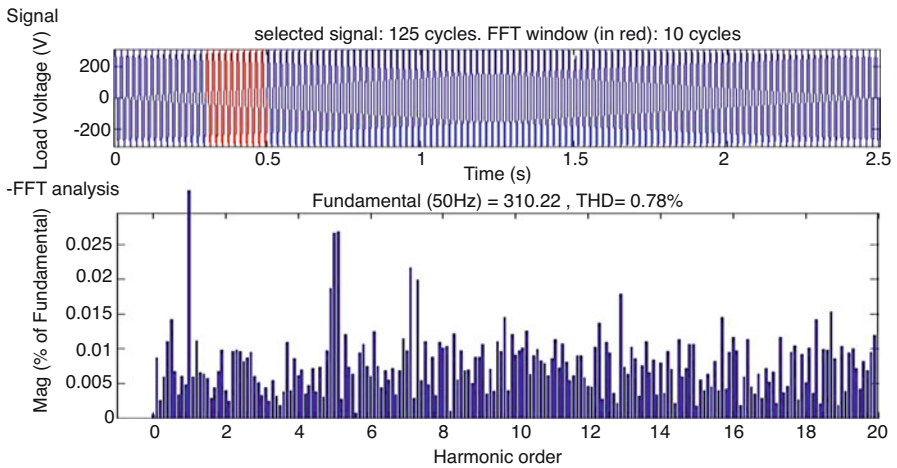


Fig. 25.6 Proposed system harmonic distortion

References

1. Liang X (2017) Emerging power quality challenges due to integration of renewable energy sources. *IEEE Trans Ind Appl* 53(2):855–866
2. Sorkhabi SS, Bakhshai A (2016) Microgrid control system based on an adaptive notch filter power processor. In: *IEEE 7th international symposium on power electronics for distributed generation systems*. <https://doi.org/10.1109/PEDG.2016.7527025>
3. Cho W, Lee W, Di Han, Sarlioglu B (2018) New configuration of multi-functional grid-connected inverter to improve both current-based and voltage-based power quality. *IEEE Trans Ind Appl*. <https://doi.org/10.1109/TIA.2018.2861737>
4. Kumari S, Choudhary SR, Sengupta A (2018) Design of adaptive notch filter. In: *International symposium on devices, circuits and systems*. <https://doi.org/10.1109/ISDCS.2018.8379638>
5. Bonaldo JP, Paredes HKM, Antenor Pomilio J (2016) Control of single-phase power converters connected to low-voltage distorted power systems with variable compensation objectives. *IEEE Trans Power Electron* 31(3):2039–2052
6. Chen S, Jobs GC (2013) Series and shunt active power conditioners for compensating distribution system faults. In: *Proceedings of Canadian Conference Electrical Computer Engineering*, Mar, 2013, pp 1182–1186
7. Zeng Z, Yang H, Zhao R, Cheng C (2013) Topologies and control strategies of multi-functional grid-connected inverters for power quality enhancement: a comprehensive review. *Renew Sustain Energy Rev* 24:223–270
8. Choi W, Lee W, Han D, Sarlioglu B (2016) New configuration of multi-functional grid-connected inverter to improve both current-based and voltage-based power quality. *Proc IEEE Energy Convers Congr Expos*:1–8. 2016
9. Zeng Z, Li H, Tang S, Yang H, Zhao R (2016) Multiobjective control of multi-functional grid-connected inverter for renewable energy integration and power quality service. *IET Power Electron* 9(4):761–770. <https://doi.org/10.1049/iet-pel.2015.0317>
10. Golestan S, Guerrero JM, Vasquez JC, Abusorrah AM, Al-Turki Y (2017) Modeling, tuning, and performance comparison of advanced SOGI-based FLLs. *IEEE Trans Power Electron*. <https://doi.org/10.1109/TPEL.2018.2808246>
11. Sekhara Reddy Chilipi R, Al Sayari N, Al Hosani KH, Beig AR (2017) Adaptive notch filter based multipurpose control scheme for grid-interfaced three-phase four-wire DG inverter. *IEEE Trans Ind Appl* 53:4015. <https://doi.org/10.1109/TIA.2017.2676098>
12. Chaudhary P, Rizwan M (2015) A grid synchronization method based on adaptive notch filter for SPV system with modified MPPT. *Eng Technol Int J Electr Comp Eng* 9(7)
13. Misra B, Nayak B (2018) Modified second order adaptive filter for grid synchronization and reference signal generation. *Int J Renew Energy Res* 8(1)
14. Fan Y, Zhang Q, Wang W, Wu Z (2018) Speed regulation system of a flux-modulated permanent-magnet in-wheel motor based on sliding mode control and adaptive notch filter. *IEEE Trans Energy Convers*. <https://doi.org/10.1109/TEC.2018.2859338>
15. Gilani SO, Ilyas Y, Jamil M (2018) Power line noise removal from ECG signal using notch, band stop and adaptive filters. In: *International Conference on Electronics, Information, and Communication (ICEIC)*. <https://doi.org/10.23919/ELINFOCOM.2018.8330569>
16. Abrar M, Khan WE (2018) Digital filtering techniques for power-line interference removal from ECG signals. In: *IEEE Conference on Information Communications Technology and Society (ICTAS)*. <https://doi.org/10.1109/ICTAS.2018.8368750>
17. Rufa'I NA, Zhang L, Chong B Performance analysis of adaptive notch filter active damping methods for grid-connected converters under a varying grid impedance. In: *2017 IEEE Manchester PowerTech 12th IEEE PES PowerTech Conference*, p 2017. <https://doi.org/10.1109/PTC.2017.7981203>

18. Chaochao J, Yixin S, Huajun Z, Shilin L (2016) Power system frequency estimation based on adaptive notch filter. In: International Conference on Industrial Informatics. <https://doi.org/10.1109/ICIICII.2016.92>
19. Wang Y, Zheng Q, Zhang H, Miao L (2018) Adaptive control and predictive control for torsional vibration suppression in helicopter/engine system. IEEE Access 6. <https://doi.org/10.1109/ACCESS.2018.2829723>

Chapter 26

Optimal DAU Placement for Smart Distribution Grid Communication Network



S. Premkumar , M. Susithra , and V. Saminadan 

Abstract The reliability and stability of communication systems are crucial for utility centres to deliver power to the consumers in an efficient manner. This chapter investigates optimal placement of Data Aggregation Units (DAUs) in smart distribution grids equipped with smart metres and smart automation devices (SADs). The DAUs are used as relay points to transmit the data from smart metres and SADs to the control centre in a cost-efficient manner. The optimization of DAU placement is done based on the iterative K-means clustering method. This work presents an optimal placement of DAUs in a wireless network, which includes smart metres and SADs that help the utilities communicate within themselves with less delay.

Keywords Data aggregation unit (DAU) · Smart automation device (SAD) · Smart grid device (SGD) · Worldwide interoperability for microwave access (WiMAX)

Abbreviations

CBR	Constant bit rate
CC	Control centre
DAU	Data aggregation unit
ertPS	Extended real-time polling service
ETE	End to end
HAN	Home area network
NAN	Neighbourhood area network
nrtPS	Non-real-time polling service
rtPS	Real-time polling service
SAD	Smart automation device
SGD	Smart grid device

S. Premkumar (✉) · M. Susithra · V. Saminadan
Pondicherry Engineering College, Pondicherry, India

- UGS Unsolicited grant service
- WAN Wide area network
- WiMAX Worldwide interoperability for microwave access

26.1 Introduction

A smart grid provides two-way communication between the service provider and the consumers, which generally consists of home area network (HAN), neighbourhood area network (NAN) and wide area network (WAN). The monitoring of the power grid status (e.g. AMI information, smart automation devices information) is gathered by DAUs, as shown in Fig. 26.1. The massive data generated from these devices are very crucial for the utilities to stabilize the grid operation. The data from these devices must not have huge delay and losses. The DAUs act as concentrators to collect the data from a number of associated smart devices at regular intervals to a central control centre (CC). The optimal placement of DAUs plays a key role in the supervision of a smart distribution grid and support of various applications. Mahsa et al. [1] present an optimization problem to find the best location for aggregator placement in a hybrid wireless and wired communication network including fibre

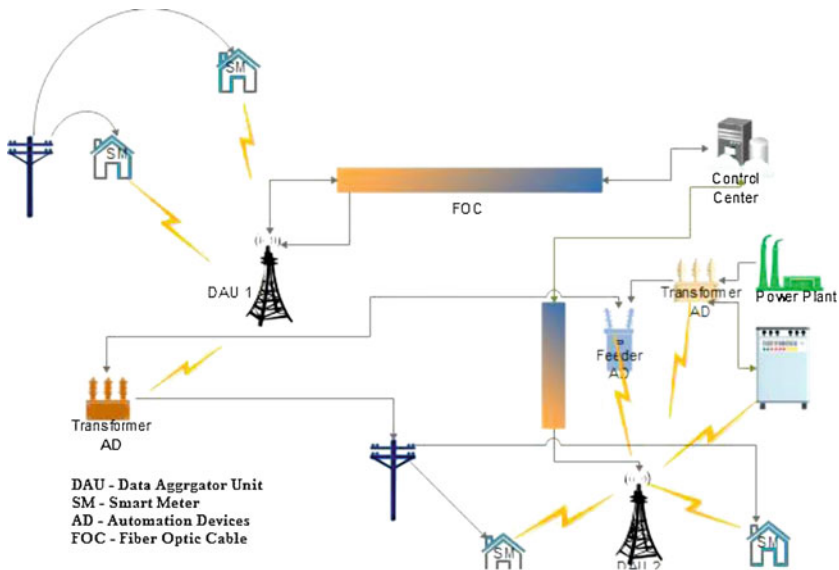


Fig. 26.1 DAU placement model in WiMAX based rural area for smart grid

optic and WiMAX (Worldwide Interoperability for Microwave Access). Souza et al. [2] propose a methodology in a zigbee-based mesh network to find concentrator placement for optimizing the performance of the network in an advanced metering infrastructure (AMI) environment. In Ref. [3], the optimal placement of DAUs for an AMI is analysed with an objective to minimize the installation and transmission cost, by only considering different metre densities. In Ref. [4], the authors propose an optimization problem to minimize the total cost of aggregator placement in a NAN environment. They also considered geographical deployment methods for the placement of gateway aggregate points. In Ref. [5], the authors designed clustered substation equipment monitoring for smart grid networks based on a wireless sensor network. Using an OPNET modeler, the QoS parameters were analysed for a clustered network, which shows better performance than a tree and mesh network.

We propose a self-sustaining wireless wide area network for smart grid and an optimization problem to achieve the minimum number of DAUs without compromising the reliability of network performance. The method proposed uses MatLab to read the location of smart metres and smart automation devices as input data. The read data then gets compiled into a network topology consisting of smart metre nodes and power line links between them.

For easier comparison, the network is then clustered into groups. In order to establish connection among the different components of a smart grid, various wireless technologies such as WiMAX, WiFi and Zigbee can be employed. The selection of an appropriate communication technology depends on a number of criteria, with the required network coverage and the types of data traffic with their quality-of-service (QoS) requirements being among the most important ones. These requirements are specific to the smart grid domain.

26.2 Clustering Smart Devices

To investigate the communication capabilities of a smart grid, the contiguous environment conditions must be taken into consideration. Here, the smart metre and the smart automation devices communicate in a point-to-point manner with a nearby DAU, which in turn forwards the data towards the CC where a virtual data concentrator is operating. While clustering the nodes of smart distribution grid K-means, it is considered to cluster the nodes of a smart distribution grid into different groups. K-means finds the optimal position of the centroid for the placement of DAU in the cluster group containing smart metres and smart automation devices. These SADs share the communication medium with other devices present in the vicinity. The distance from the SADs to the DAUs and the cluster density are analysed by means of improved K-means algorithm.

26.3 System Model

The IEEE 802.16 WiMAX standard is seen as a versatile wireless network that could efficiently handle rural areas with broad coverage. It employs five different service classes to handle applications with different QoS associated with uplink and downlink scheduling that could be adapted for smart grid applications.

The Unsolicited Grant Service (UGS) is used to reduce resource assignment overhead to serve the constant bit rate (CBR) of a fixed packet size periodically. Real-time Polling Service (rtPS) is used to minimize variable traffic overhead in delay-sensitive applications that do not need periodic resource allocations. The extended real-time Polling Service (ertPS) class is used to meet variable data rates with periodic allocations for real-time traffic such as VoIP service. The non-real-time Polling Service (nrtPS) is designed to provide a minimum QoS support for non-real-time applications with low data traffic rates. The BE service does not support delay constraints and offers no guarantee to data traffic. Table 26.1 shows the traffic classes and the communication requirements of devices used in a smart grid [6, 7]. Smart metres and smart automation devices are devices that are installed on customer premises and power system equipment. These devices regularly report the real-time status of this equipment to the utilities for automated monitoring to take necessary controls. The different types of automation devices that are included in the simulation are monitoring, situational awareness, protection and control, and smart metres. The simulation scenario that is considered for our simulation is a rural environment. In this case, WiMAX technology based smart devices are used as transceivers.

Table 26.1 Traffic classes and communication requirements

Use case	Direction	Packet size (bytes)	Data rate (kbps)	Active/idle	Latency	Traffic type	Proposed scheduled type
Situational Awareness	UL	256	5.0	1/5	1000	Deterministic	nrtPS
Monitoring	UL	384	300	Continuous	100	Deterministic	UGS
Control	UL	128	5.0	1/5	100	Random	rtPS
Protection	UL	192	150	Continuous	20	Random	ertPS
Smart metring	UL	256	1.0	0.1/4.0	5000	Deterministic	BE
Situational awareness	DL	256	1.0	1/5	1000	Deterministic	nrtPS
Monitoring	DL	128	10	Continuous	100	Deterministic	UGS
Control	DL	128	1.0	1/5	100	Random	rtPS
Protection	DL	192	150	Continuous	20	Random	ertPS

26.4 Improved K-Means Algorithm

The primary objective is to install a minimum number of DAUs to provide maximum coverage for all SADs. A rural area with single-hop communication to ensure low latency is considered. For this simulation, a WiMAX network scenario of 5×5 kilometres is considered. Each smart automation device in the geographical area under examination has a fixed location coordinate, that is, (SGDx, SGDy), while the DAUs are placed at location coordinates (aggx, aggy) to be identified by the proposed optimization technique. To optimize the number and location of aggregators, only the Euclidean distance between the IEDs and the aggregators in the SG environment were considered in previous studies. In this simulation, the various smart automation devices present in the smart distribution network were taken into consideration based on their QoS and their co-ordinates. All the smart automation devices will be given weights based on their QoS requirements, which means two or more smart automation devices that are within physical vicinity can be linked to different DAUs subject to their location and QoS constraint. The key objective of the proposed optimization algorithm is to find the coordinates of the DAUs and ‘ k ’ to minimize the latency of the smart automation devices.

The total average delay is the delay that occurs during the information transfer from all the smart automation devices to and from the respective DAUs to the CC. To model this, a rural area with n SADs, ‘ k ’ DAUs and one utility centre is considered. The number of smart automation devices taken into consideration for different environments is obtained from Ref. [8]. The SAD is grouped to form ‘ k ’ clusters based on the coordinates and their weights [9].

In each cluster, there are ‘ n_j ’ number of nodes with the condition

$$\sum_{j=1}^k n_j = n \quad (26.1)$$

where ‘ n ’ is the total number of SADs.

Mathematically, the DAU placement problem can be written as

$$\text{Minimize } z = \sum_{i=1}^n \sum_{j=1}^k d_{ij} * x_{ij} \quad (26.2)$$

where,

d_{ij} is the distance between the i th SAD and j th cluster $w(x_i)$ represent the weight of each node.

In the modified K-means algorithm, weight measure is included for each SAD to form a cluster. The cluster centres will be assigned such that it will be optimized. The optimal location of the cluster centre will be assigned such that the nodes with maximum weights will be given high priority.

The Euclidian distance d_{ij} between the SAD and the centroid is calculated using

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (26.3)$$

where,

$i = 1, 2, \dots, n$, $j = 1, 2, \dots, k$. x and y refers to the coordinates of the i th SAD and j refers the cluster in the Euclidian plane. The Euclidian distance d_{ij} is obtained for all i to every j .

The number of clusters is calculated based on the weights $w(x_i)$ of SADs and capacity of cluster (C) as follows.

$$K = \left\lceil \frac{\sum_{i=1}^n w(x_i)}{C} \right\rceil \quad (26.4)$$

The Euclidian distances between the SAD and all the k centroids are calculated. Priority values for each SAD are calculated to find the appropriate centroid based on equation

$$\text{Priority } P_i = \frac{d_{ij}}{w_i} \quad (26.5)$$

The centroid (x_i, y_i) for each cluster is calculated based on the membership function of each SAD.

$$m(c_j|x_i) = \frac{\|x_i - c_i\|^{-p-2}}{\sum_{j=1:k} \|x_i - c_i\|^{-p-2}} \quad (26.6)$$

$$c_j = \frac{\sum_{i=1:n} m(c_j|x_i) w(x_i) x_i}{\sum_{i=1:n} m(c_j|x_i) w(x_i)} \quad (26.7)$$

The improved K-means algorithm has been adapted to find optimal DAU placement. It considers priorities for the different types of nodes based on distance and demand. The nodes with high demand and shorter distance will be put into the initial cluster and the nodes with less demand and longer distance will be put into the other clusters.

Steps to find optimal placement of DAUs using improved K-means algorithm are as follows [10]:

Step 1: Select the number of cluster group k .

Step 2: Randomly create k cluster groups and define the cluster centre, or generate k random points as cluster centres.

Step 3: Based on the priority, weight and the Euclidian distance the centroid has to be located for each cluster.

Step 4: Allocate each point to the adjacent cluster centre.

Step 5: Recompute the original cluster centres.

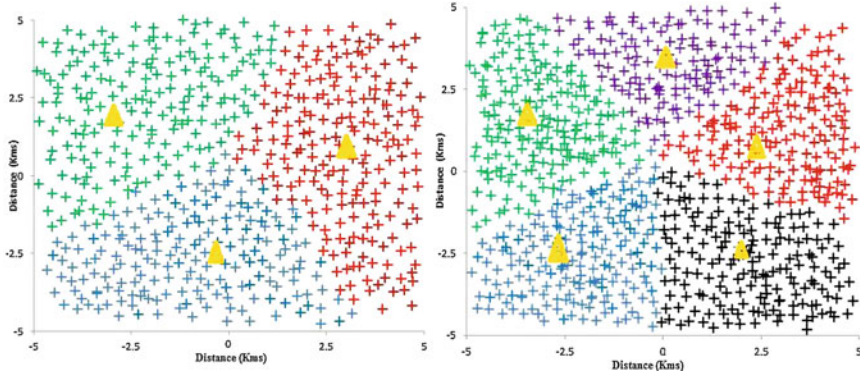


Fig. 26.2 DAU placement using improved K-means algorithm for ($k = 3$) and ($k = 5$)

Step 6: The two previous steps are to be repeated until the convergence criterion is attained.

Step 7: A priority measure to select the nodes for cluster assignment to the nearest cluster based on high demand and short distance is considered.

Figure 26.2 shows the results of the optimal placement of the DAU locations using improved K-means algorithm for $k = 3$ and $k = 5$ for the rural environment. The yellow triangles represent the candidate sites where the DAUs can be installed. When $k = 3$, an average of 100 devices are connected to one DAU, and for $k = 5$, an average of 60 devices are connected to one DAU. The number of DAUs required to connect all the devices in a rural environment gets reduced to a minimum due to the implementation of improved k-means algorithm.

The improved k-means performance characteristics are examined for all the devices considered in the simulation based on the WiMAX network. The simulation is done in OPNET modeler 14.5. For performance evaluation, two different scenarios are developed. The first scenario is modelled for $k = 3$ based on the optimal position of the DAUs obtained from the proposed improved K-means algorithm. The smart grid devices are placed according to the coordinates in the WiMAX infrastructural mode with a centre frequency of 1800 MHz. The topology dimension for this simulation is 5 km \times 5 km, supporting up to 1000 smart grid distributed devices. The DAUs are connected to the CC through a PPP_SONET_OC3 link. The smart grid devices gather information and transmit the data over the WiMAX link and the services will be provided according to the weights, priorities and WiMAX service classes assigned to the devices. The smart grid devices are connected to the base stations assigned according to the clustering process by specifying the MAC address in the base station and smart grid devices. The proposed improved K-means clustering deployment scheme selects the appropriate placement locations for the BSs from the candidate positions by considering the weights and priorities. The latency decreases between the SGD and the CC after the placement of BS using the proposed improved K-means algorithm.

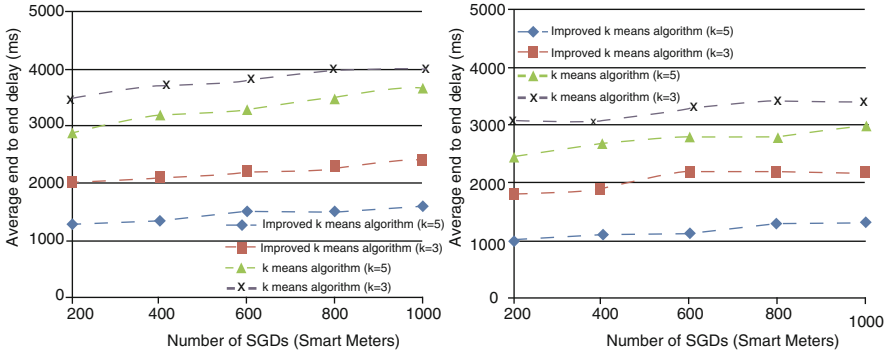


Fig. 26.3 Uplink and downlink ETE delay for smart metres

Table 26.2 Mean and standard deviation of optimal DAU placement

Number of SGDs	K-means without priority		Improved K-means with priority	
	Mean	Standard deviation	Mean	Standard deviation
For 500 SGDs with $K = 3$	15.671	4.295	14.823	3.527
For 500 SGDs with $k = 5$	12.483	3.142	11.164	2.279

In the K-means algorithm, no specific criterion was set in selecting the initial centroid because the SGD with least demands is considered for clustering even before the SGD with highest demands. The improved K-means considers both the priorities and the weights of all the SGDs for clustering to find the optimal centroid in which the base station can be placed. Figure 26.3 shows that the End to End (ETE) delay for smart metre devices considered for the smart distribution grid in a rural environment is significantly reduced both for uplink and downlink using improved K-means algorithm.

Table 26.2 compares the mean and standard deviation of K-means and improved K-means algorithm for optimal DAU placement. The average demand of each SGD is calculated by finding the mean of all the clusters. The standard deviation indicates the measure by which the SGDs are clustered in a balanced manner. With improved K-means algorithm for $k = 5$, the mean value decreases from 12.483 to 11.164 and the standard deviation decreases from 3.142 to 2.279. Similarly, for $k = 3$, the mean value decreases from 15.671 to 14.823 and the standard deviation decreases from 4.295 to 3.527 as compared with K-means algorithm. The results show that the improved K-means algorithm based SDGN has continuous connectivity with low delay, packet drop ratio and high throughput.

26.5 Conclusion

The DAU placement planning plays an important role in the stability of smart grid communication. DAU placement is formulated as an integer program. The improved K-means algorithm is adapted to obtain optimal solutions for the placement of

DAUs. Simulation results show that fewer DAUs are required for the aggregation of data from the devices. The better placement of DAUs also achieves improved performance in terms of delay and throughput when compared with the conventional K-means algorithm.

References

1. Tavasoli M, Yaghmaee MH, Mohajerzadeh AH (2016) Optimal placement of data aggregators in smart grid on hybrid wireless and wired communication. In *Smart ENERGY Grid Engineering (SEGE)*, pp 332–336
2. Souza G, Mestrando FV, Lima C, Junior G, Castro M, Sérgio A (2013) Optimal positioning of GPRS concentrators for minimizing node hops in smart grids considering routing in mesh networks. In *Innovative Smart Grid Technologies Latin America (ISGT LA)*, 2013 IEEE PES conference on, IEEE, pp 1–7
3. Aalamifar F, Shirazi GN, Noori M, Lampe L Cost-efficient data aggregation point placement for advanced metering infrastructure. In *Smart Grid Communications 2014 IEEE International Conference on*, pp 344–349
4. Eason G, Noble B, Sneddon IN (1955) On certain integrals of Lipschitz-Hankel type involving products of Bessel functions. *Philos Trans R Soc Lond A* 247(935):529–551
5. Jie H, Kai M, Yang Y Research on wireless network topology of substation equipment monitoring based on QoS. In *Intelligent Computation Technology and Automation (ICICTA)*, 2014 7th international conference on, pp 605–608
6. Aalamifar F, Lampe L (2017) Optimized WiMAX profile configuration for smart grid communications. *IEEE Trans Smart Grid* 8(6):2723–2732
7. Castellanos G (2012) Wireless communication network architecture for the smart grid applications (Doctoral dissertation, thesis, University of Newcastle Callaghan, NSW 2308 Australia)
8. Huang X, Wang S (2015) Aggregation points planning in Smart Grid communication system. *IEEE Commun Lett* 19(8):1315–1318
9. Kerdprasop K, Kerdprasop N, Sattayatham P (2005) Weighted k-means for density-biased clustering. In *International Conference on Data Warehousing and Knowledge Discovery*, Springer, Berlin, Heidelberg, pp 488–497
10. Geetha S, Poonthalir G, Vanathi PT (2009) Improved k-means algorithm for capacitated clustering problem. *INFOCOMP* 8(4):52–59

Chapter 27

Long-Term Forecasting of Hybrid Renewable Energy Potential Using Weibull Distribution Method in Coimbatore



Anuradha J, Soundarrajan A, and Rajan Singaravel M M

Abstract Assessment of hybrid renewable energy is a critical factor for the suitable development of distributed generators (DG). Sizing of DGs needs accurate forecasting. This chapter presents a forecast of wind speed and solar irradiance using the two-parameter Weibull Distribution (WD) method. The dimensionless shape parameter ‘k’ and scale parameter ‘c’ are determined based on hourly global solar irradiance and wind speed from time series data during 2004–13 is used to estimate the Weibull parameters: hourly global solar irradiance and wind speed predicted all around a year. The performance of the Weibull Distribution method is analysed using Mean Absolute Error (MAE) and Mean Squared Error (MSE). The obtained result indicates the EPF method is suitable for prediction of mean hourly solar irradiance and wind speed.

Keywords Weibull distribution · Probability distribution function · Energy Pattern Factor Method · Wind speed · Solar irradiance

Abbreviations

CEA	Central Electricity Authority
DG	Distributed Generators
EPF	Energy Pattern Factor method
HRES	Hybrid Renewable Energy Sources

Anuradha J (✉) · Soundarrajan A
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India

Rajan Singaravel M M
Department of Electrical and Electronics Engineering, NIT, Puducherry, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_27

MAE	Mean Absolute Error
MSE	Mean Squared Error
WD	Weibull Distribution

27.1 Introduction

For building a sustainable future, renewable energy resources play a major role in the generation of electrical energy. India is the world's third-largest electricity producer and fourth-largest consumer in the world. The Government of India is focusing on the renewable energy sector to promote the industrial sector in an environment-friendly way. India continues to rely heavily on thermal power. Out of the 344 GW of installed capacity in the country as of June 2018, 57% comes from thermal units, which use coal as fuel. While the intension of the Central Electricity Authority (CEA) is to reduce Energy Pattern Factor method (EPF) the fossil fuel dependence to below 50% by year 2022.

Efficient use of renewable energy is the fastest, cheapest and spotless way to meet the rising energy demand and power the nation's growth. Hybrid renewable energy sources such as wind and solar photovoltaic (PV) have been taken into consideration. Both the sources are intermittent in nature and also complementary. Integrated resources give the assured power rather than one. Integration of wind and solar PV is necessary for getting the promised output power and stability of the power system. Objective of this work is the assessment of hybrid renewable energy sources (HRES) in Coimbatore using the two-parameter Weibull Distribution method.

To ensure secure and economic growth of solar PV and wind energy into the existing grid, accurate forecasting of HRES is critical [1, 2]. Seasonal and climate conditions impact wind speed and solar irradiation. Recently many researchers focused on the area of forecasting of wind and solar energy [3]. Output power of a wind generator is proportional to the cube of the wind speed and solar PV depends on irradiance. Hence, accurate wind speed prediction is essential and more important for wind resource assessment than solar power.

A variety of methods are proposed for prediction of wind power. Time series data on wind speed are chaotic, and prediction accuracy depends heavily on the time frame. The accuracy of prediction decreases with increase in the prediction time frame [4]. Wind speed forecasting techniques are categorized into five methods: (a) physical method, (b) statistical method, (c) artificial methods, (d) spatial correlation methods, and (e) hybrid methods. Each and every method has its own characteristics.

Objective of this work is the long-term assessment of HRES in Coimbatore using the two-parameter WD method. It helps in the planning to maximize the generation of electric power from renewable energy sources such as wind and solar.

27.2 Renewable Energy Potential of Coimbatore

27.2.1 Scenario of Coimbatore

Coimbatore located at the foot of the Western Ghats, is well known as the Textile City with established industries. It is the third-largest city in Tamil Nadu with a population of more than 15 lakh. This city has well-developed educational institutions, hospitals and machine tool industries. It has been selected by the Government of India to implement the Smart City project.

As part of the Smart City project, uninterrupted power should be provided for all. Electric power consumption is increasing drastically due to increase in industrial activities and economic development. Apart from power produced from fossil fuels, green energy sources such as wind and solar energy make the city green and clean. In order to supply for the peak energy demand, a combination of wind and solar PV gets more attention. Figure 27.1 indicates India’s target to implement renewable energies totalling 175 GW by the year 2022.

27.2.2 Characteristics of Wind

The motion of air masses is called wind. Air masses move because of uneven heating of the earth. The causes of wind can be global as well as local and regional phenomena. Wind is characterized by speed and direction. Wind speed varies continuously as a function of time and height. Wind turbines use wind energy close to the ground. Wind speed forecasting plays a major role in balancing the power system. The power of an air mass that flows at the speed of ‘v’ [5].

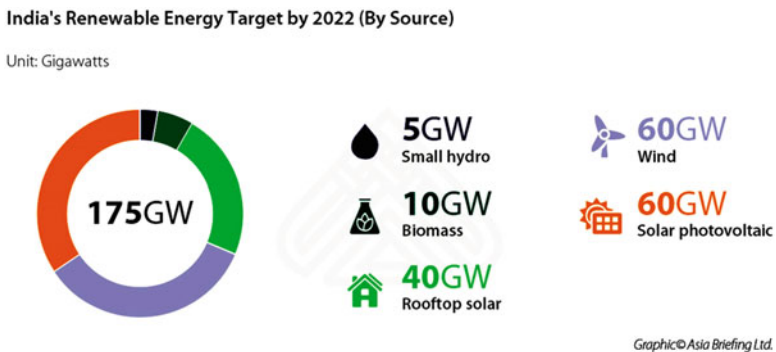
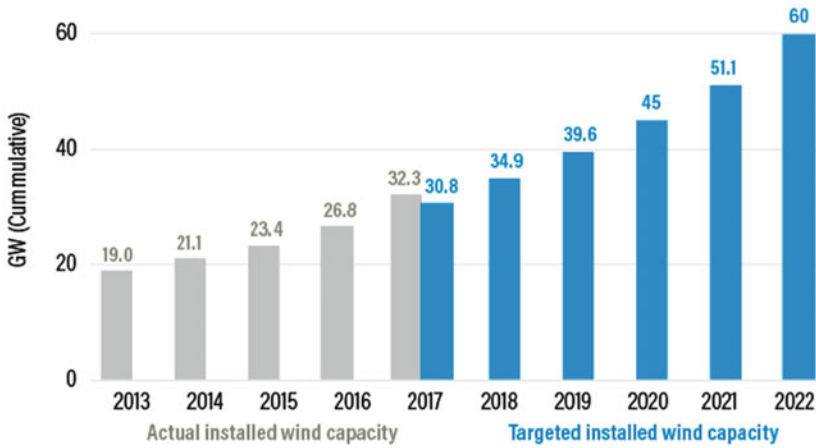


Fig. 27.1 India’s renewable energy target by the year 2022

India's Year-on-Year Targets to Reach 60 GW Wind Goal



Source: India's Ministry of New and Renewable Energy (MNRE)
 Note: All years in the chart are fiscal years: 1 April–31 March



Fig. 27.2 India's wind goal in the year 2022. (Source: Blooming New Energy Finance (BNEF). The Economic Times)

$$\text{Power in wind} = \frac{1}{2} \rho A v^3 \text{ in (watts)} \tag{27.1}$$

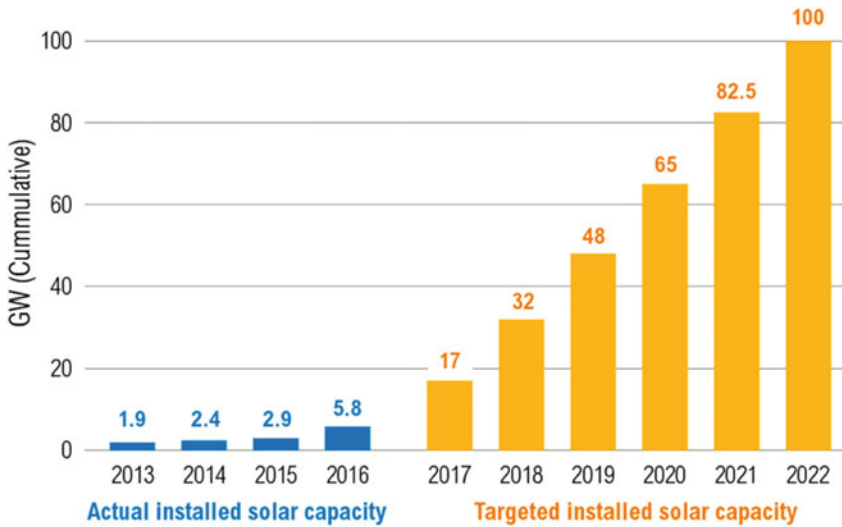
Where ρ - Air density (kgm^{-3}), v - Wind speed (ms^{-1}) A-Intercepting area.

Power in the wind is proportional to air density, the intercepting area A and the cube of the velocity v. Air density is a function of air pressure and air temperature, which both are above the ground level. Wind speed is a random parameter; usually statistical methods were applied for analysis the historical wind speed. Figure 27.2 gives a detailed picture of wind energy penetration in the Indian power system.

27.2.3 Characteristics of Solar Energy

The output of solar forecasting is solar irradiance and PV power. Solar forecasting is not widely used. Prediction of solar power is used in the initial stages only [6]. The conversion efficiency is less (25–30%). An efficient use of the varying energy output of photovoltaic (PV) systems needs reliable forecast statistics. Technological development reduces the cost of solar cells. Researchers are focused on the development of solar energy. Figure 27.3 shows that focus on solar energy is increasing.

India Sets Year-on-Year Targets to Reach Ambitious 2022 Solar Goal



Notes: FY = All years in chart are fiscal year from April 1 to March 31; 1 GW = 1,000 MW.
 Sources: Bloomberg New Energy Finance (BNEF); The Economic Times.



Fig. 27.3 India’s solar goal in the year 2022. (Source: Blooming New Energy Finance (BNEF). The Economic Times)

27.3 Weibull Distribution Method

The Weibull model is used with time series data. A statistical time series analysis is the representation of time series. To predict the future values from the distribution of past wind speed data values, the two-parameter Weibull Distribution method is used [6]. It represents various distribution characteristics, where shape parameter and scale parameter are changed.

Weibull distribution is the perfect fit for wind speed distribution. In many research studies, Weibull distribution has been considered as a useful tool for wind energy estimation by expressing annual mean wind speed variations. Weibull PDF is described by scale and shape parameters. The wind speed distribution function is characterized by

$$f(v) = \frac{k}{c} \left(\frac{v}{c}\right)^{k-1} \exp \left[-\left(\frac{v}{c}\right)^k \right] \quad (27.2)$$

$$f(q) = \frac{k}{c} \left(\frac{q}{c}\right)^{k-1} \exp \left[-\left(\frac{q}{c}\right)^k \right] \quad (27.3)$$

Where $f(v)$ is the probability of occurrence of wind speed $v(v \geq 0)$; q is irradiation (W/m^2) k is the shape parameter, c is the scale parameter. The complementary cumulative Weibull, distribution function $F(v)$ is the integral of the PDF is given in Eq. (27.3)

$$F(v) = 1 - \exp \left[-\left(\frac{v}{c}\right)^k \right] \quad (27.4)$$

Using the inverse transform method, the probable wind speed is given in Eq. (27.4)

$$v = c \left[-\ln (1 - U) \right]^{\frac{1}{k}} \quad (27.5)$$

Where, U is the uniformly distributed random variable between $[0, 1]$. The equation can be simplified as follows

$$v = c \left[-\ln (U) \right]^{\frac{1}{k}} \quad (27.6)$$

27.3.1 Energy Pattern Factor Method

The two parameters of the Weibull PDF are computed using the EPF method. This method is associated with the average data of wind speed/solar irradiance [6, 7] and is defined by Eq. (27.7).

$$E_{pf} = \frac{(v^3)_m}{(v_m)^3} = \frac{\left(\frac{1}{n} \sum_{i=1}^n v_i^3\right)}{\left(\frac{1}{n} \sum_{i=1}^n v_i\right)^3} \quad (27.7)$$

Where, E_{pf} is the energy pattern factor, v_m is mean wind speed, m/s; v_i hourly wind speed m/s; i number of measured hours; once the E_{pf} is calculated then shape and scale parameter are estimated using Eqs. (27.8 and 27.9).

$$k = 1 + \left(3.69 / (E_{pf})^2\right) \quad (27.8)$$

$$c = \frac{v_m}{\Gamma.(1 + 1/k)} \quad (27.9)$$

Where, Γ is the gamma function. For wind speed forecasting, v represents the hourly mean wind speed, then wind speed is replaced by solar irradiance q for forecasting of solar irradiance [8].

27.3.2 Description of Data Set

Latitude and longitude of Peelamedu, Coimbatore, is **11.0332° N and 77.0277° E**. Its elevation is 430 m above the sea level. Hourly mean solar irradiance and wind speed data are collected from NSRDB data viewer.

27.4 Result and Discussions

27.4.1 Wind Speed Prediction

The data from NSRDB data viewer from 2004 to 2013 is collected in time series and pre-processed [9]. Based on historical data, the shape and scale parameters of WD are calculated using the EPF method. Then the probability of the next year's wind speed is predicted.

The parameters for WD using the EPF method are calculated based on geographical data. Long-term wind speed is forecast. The month-wise probability distribution of wind speed at 10-m height is plotted (Fig. 27.4).

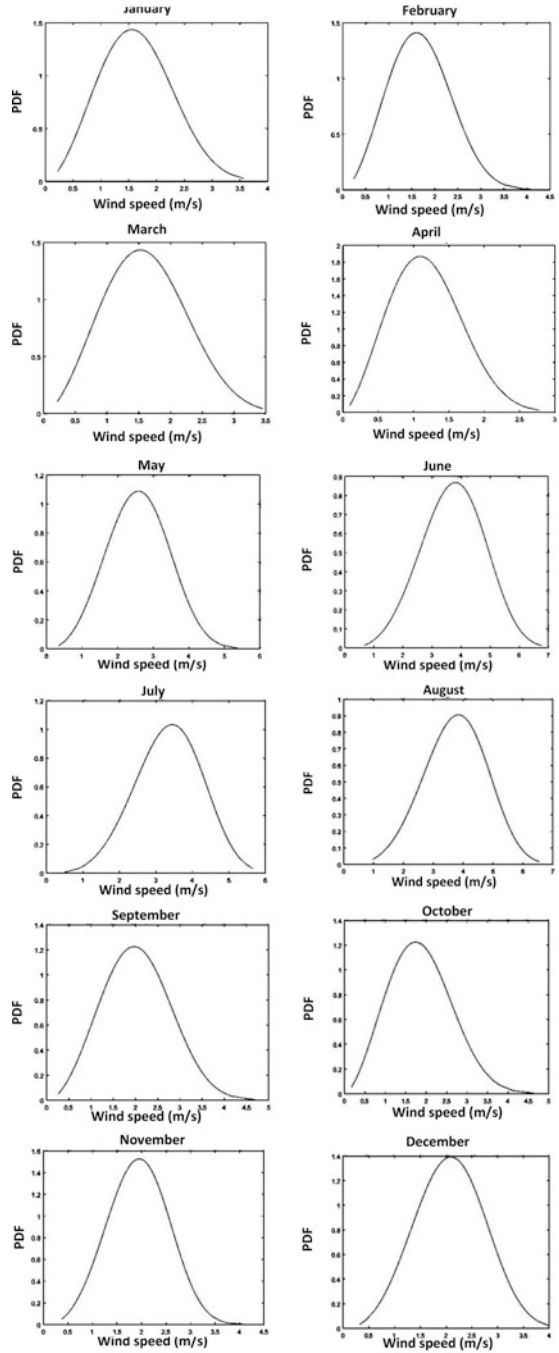
27.4.2 Prediction of Solar Irradiance

From the data set, solar irradiance for day hours 6.30 am–6.30 pm is taken into account. The average hourly global solar irradiance time series is framed. Probability distribution of hourly mean solar irradiance is predicted from shape and scale parameters of WD. The average per day global horizontal surface irradiance is also calculated. The average per day global irradiance for the year 2014 is shown in Fig. 27.5.

The performance analysis for the forecasting method is evaluated using Mean Absolute Error (MAE) and Mean Square Error (MSE) methods. The smaller the values, the better the forecasting performance. The following table gives prediction analysis for the year 2014 (Table 27.1).

Thus, the result shows that the region has more potential of solar and wind energy. The hybrid renewable energy sources surely give promised output power.

Fig. 27.4 Probability distribution of wind speed at Coimbatore



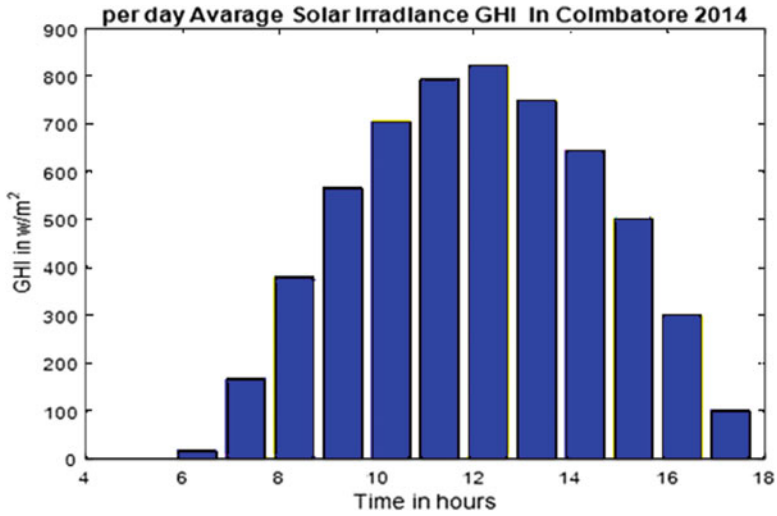


Fig. 27.5 Average per day solar irradiance (watts/m²)

Table 27.1 Prediction analysis for the year 2014

Month	Shape factor(k) (k)	Scale factor(c)c)	MSE(Mean M MSE	MA (MAEE
January	2.7067	0.7507	0.1517	0.3103
February	2.7366	0.7711	0.1580	0.3219
March	2.6614	0.7411	0.1416	0.3048
April	2.5217	0.5334	0.0870	0.2500
May	3.3040	1.7421	0.2940	0.4300
June	3.8184	1.6800	0.3403	0.4644
July	4.0850	1.5031	0.2881	0.4237
August	3.8972	1.8030	0.3757	0.4198
September	2.8900	0.9291	0.3029	0.4465
October	2.6090	0.8533	0.2251	0.3908
November	3.4754	0.8779	0.1684	0.3027
December	3.4030	0.9435	0.1466	0.3202

27.5 Conclusion

This chapter explores the assessment of HRES in a simple and convenient way. The errors proved the EPF method is efficient in the prediction of wind speed and solar irradiance. From historical data, the yearly mean wind power density varies with a standard deviation of 10% more or 10% less. The synaptic and diurnal peaks may affect the long-term power balancing of the power system. Even though the EPF method fails to predict the synaptic peak, it is largely comparable in long-term

prediction accuracy. It also offers a simpler calculation in the prediction of both wind speed and solar irradiance. This prediction helps traders of wind and solar power plants to look forward to greater economic profits.

References

1. Nema P, Nema RK, Rangnekar S (2009) A current and future state of art development of hybrid energy system using wind and PV-solar: a review. *Renew Sust Energy Rev* 13:2096–2103
2. Ehsan A, Yang Q (2018) Optimal integration and planning of renewable distributed generation in the power distribution networks: a review of analytical techniques. *Appl Energy* 210:44–59
3. Al-Masri HM, Ehsani M (2016) Feasibility investigation of a hybrid on-grid wind photovoltaic retrofitting system. *IEEE Trans Ind Appl* 52:1979–1988
4. Azad HB, Mekhilef S, Ganapathy VG (2014) Long – Term wind speed forecasting and general Pattern recognition using neural networks. *IEEE Trans Sust Energy* 5(2):546–553
5. Vafaeipour M, Valizadeh MH, Rahbari O, Eshkalag MK (2014) Statistical analysis of wind and solar energy potential in Tehran. *Int J Renew Energy Res* 4(1):233–239
6. David A, Ngwa NR (2013) Global solar radiation of some regions of Cameroon using the linear angstrom and non-linear polynomial relations (part I) model development. *Int J Renew Energy Res (IJRER)* 3:984–992
7. Kidmo DK, Danwe R, Doka SY, Djongyang N (2015) Statistical analysis of wind speed distribution based on six Weibull methods for wind power evaluation in Garoua, Cameroon. *Rev Energ Renouv* 18:105–125
8. Rajan Singaravel MM, Arul Daniel S (2015) Sizing of hybrid PMSG-PV system for battery charging of electric vehicles. *Front Energy* 9(1):68–74
9. Kadhem AA, Wahab NIA, Aris I, Jasni J, Abdalla AN (2017) Advanced wind speed prediction model based on a combination of Weibull distribution and an artificial neural network. *Energies* 10:1744–1760

Chapter 28

Efficient and Improved ANN-Based Voltage-Rise Mitigation Strategy in Distribution Network with Distributed Solar Photovoltaic System



Neenu Thomas, R. Jayabarathi, and T. N. P. Nambiar

Abstract As the cost of photovoltaic (PV) generation systems reduces, more consumers will add grid-tied roof-top PV systems to low-voltage (LV) distribution networks in a widespread manner. Transmission-line impedance and load variations will influence the power transfer capability and voltage profile of the system. This chapter describes the influence of line impedance and load change on the voltage profile in the distribution network. An analysis on the voltage profile due to the variation of load change and line impedance is investigated in a representative LV distribution network and a new control strategy-based on Artificial Neural Network (ANN) is introduced for mitigating the voltage-rise issue in the system. The system is simulated using a MATLAB/SIMULINK platform.

Keywords Solar PV systems · Line impedance · Load change · Distribution network · artificial neural network · Voltage rise mitigation

Abbreviations

LV	Low Voltage
ANN	Artificial Neural Network
MPPT controller	Maximum Power Point Tracking controller
LMBP	Levenberg–Marquardt Back Propagation

Neenu Thomas (✉) · R. Jayabarathi · T. N. P. Nambiar
Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India

28.1 Introduction

Roof-top grid-tied solar photovoltaic (PV) systems are installed by consumers along the distribution line. Distributed PV systems will be able to meet the power demand as well as export power to the grid. Distribution networks generally have large Resistance–Reactance ratios (R/X), which affects the power transfer capability in the system [1–2]. Although the high penetration of PV systems has advantages such as reduced energy losses and efficient voltage management [3–4], the distribution system experiences a set of drawbacks. As power generation is highly dependent on solar energy, the output power of PV generation fluctuates [5]. When power generation is higher and local demand in the distribution networks is lower, power might flow from the PV systems to the domestic load. This leads to a surplus amount of power flowback to the grid. This reverse flow of power results in serious issues such as voltage rise on the distribution network, fluctuation in the system’s frequency and protection relay’s desensitization [6–8]. Among these, the over-voltage phenomenon is one of the critical issues in distribution systems. So, the mitigation strategy for the voltage-rise issue must be addressed when load change happens in the highly penetrated LV network [9].

For a stable distribution system, maintaining a permitted voltage profile along the line is very crucial. By adjusting the tap-changing transformer, the amount of voltage rise can be estimated in the LV network. Some authors suggested techniques for mitigating those negative impacts, such as storage device approach, reactive power compensation using flexible alternating current transmission system (FACTS) devices, static synchronous compensator (STATCOM), static VAR compensator (SVC), reactive power control and active power reduction approach [10–14].

The main aim of this chapter is to analyse the control strategy based on a PID controller and ANN-based controller to prevent the over-voltage issue for different load changes in the LV distribution network. The chapter is summarized as follows. Sections 28.2 and 28.3 describe the proposed control algorithm for the system. Simulation results with conventional PID approach and the proposed ANN-based approach have been presented in Sect. 28.4. Conclusions are included in Sect. 28.5.

28.2 System Model

Integration of solar PV systems into distribution networks is very promising nowadays. To identify the issues in the integration of PV systems into the LV grid, a typical distribution network is considered. The distribution line with three nodes having R/X ratio of 1 has been taken into consideration for analysis. Three nodes, each separated by a line impedance of $1 + j1$ ohm, are considered as different cases for the study. Even though this is not a realistic network to examine the effect of impedance on the voltage profile, a preliminary study is done on this network. Based upon the result of this study, a realistic network will be considered in future.

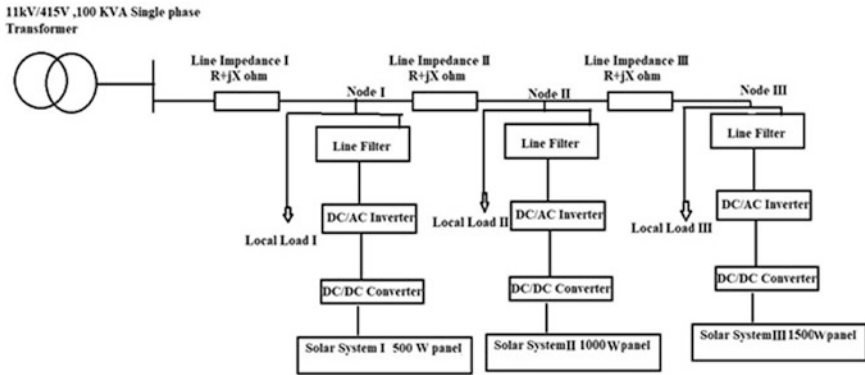


Fig. 28.1 Block diagram of the proposed system

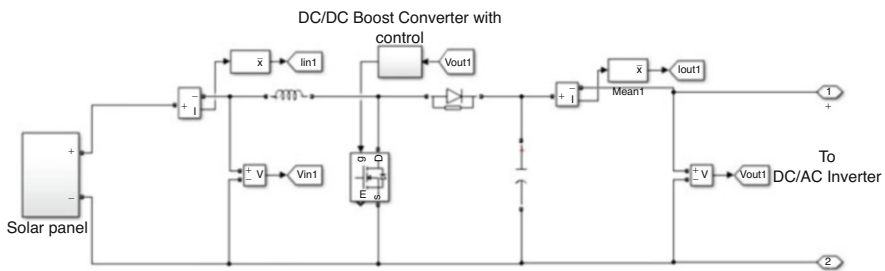


Fig. 28.2 Modelling of system with solar panel and DC/DC boost converter

The system considered for study is depicted in Fig. 28.1. Each node supports solar PV generators of 500 W, 1000 W and 1500 W along the distribution line, together with a local static resistive load of 1000 W. Each system on consumer premises consists of a solar panel and a DC/DC boost converter with a Maximum Power Point Tracking (MPPT) controller. The DC/DC boost converter output is connected to the utility grid via a DC/AC inverter along with an LCLR filter. The MPPT charge controller is connected to tackle the maximum power from PV panels using the Perturb and Observe algorithm. DC/AC inverter control is implemented by $\alpha\beta$ to dq transformation with a Phase Locked Loop (PLL).

As shown in Fig. 28.1, solar panels of different ratings are mathematically modelled and a DC/DC converter is used for the boosting of the voltage from each solar panel. The output obtained from the DC/DC boost converter is given to the DC/AC H bridge inverter. The LCLR filter is used for reducing the harmonic content in the output of the inverter. Synchronization between the grid and PV systems is done with the help of a Phase Locked Loop. Modelling of different subsystems is presented in Figs. 28.2 and 28.3. The overall system investigated is shown in Fig. 28.4. Different parameters for the system are listed in Table 28.1 and the entire system is simulated in MATLAB.

As shown in Figs. 28.2 and 28.3, inverter control logic is implemented by a conventional rotating frame such as $\alpha\beta$ to dq transformation along with a PLL. Direct

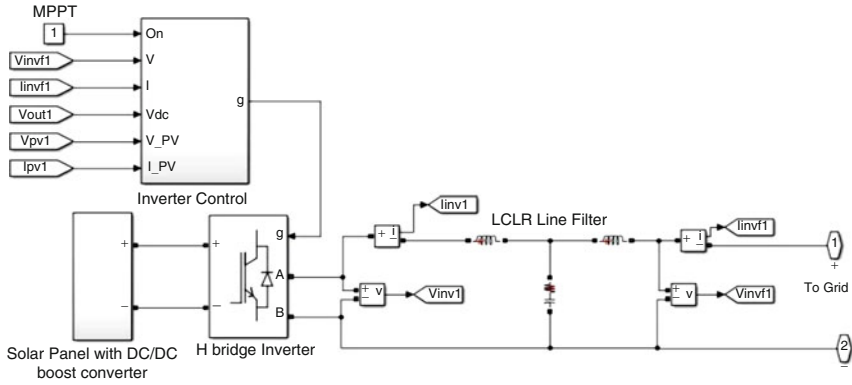


Fig. 28.3 Modelling of system containing H bridge inverter, solar panel with boost converter and line filter

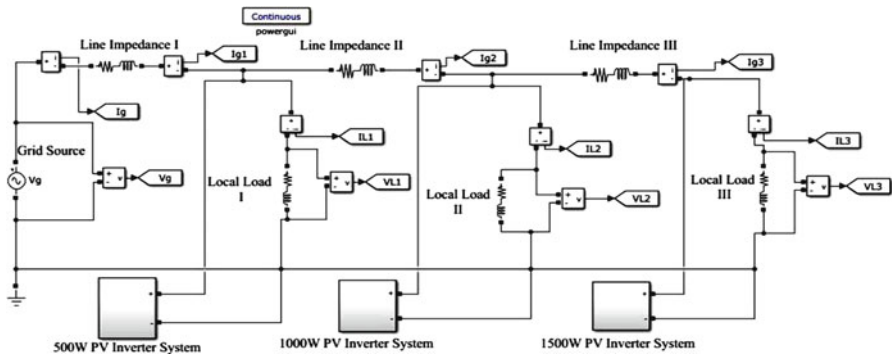


Fig. 28.4 Simulation circuit of system studied

axis and quadrature axis components of current are controlled using a discrete PID controller, which thereby generates PWM signals.

As shown in Fig. 28.4, the solar PV system, together with the inverter and line filter, is connected to the grid having distribution lines with parameters R and X values. In addition to this, local static loads are connected to each PV system. When load demand is more, both the inverter and the grid will supply power to the load. But when demand is less, the inverter itself is capable of meeting the power demand.

28.3 Proposed Control Algorithm

In the control of an inverter, the selection of the modulation index value, which is the ratio of reference voltage to the carrier voltage, is much more crucial. The objective of this work is to obtain the modulation index value in which the inverter supplies the

Table 28.1 System Description

Solar PV system	1000 W, 500 W and 1500 W
Local load	$900 + j435.88$ at each node
Sending end voltage	$230 < 0^\circ$ V
Line impedance	$1 + j1$ between each nodes (3 cases)
Boost DC/DC converter	
Inductors and capacitors	0.25 mH, 220 μ F (500 VA PV inverter)
	0.5 mH, 470 μ F (1000 VA PV inverter)
	0.1 mH, 47 μ F (1500 VA PV inverter)
LCLR filter (designed for resonant frequency = 5000 Hz)	
Inverter side inductance	16.8 mH (500 VA PV inverter)
	8 mH (1000 VA PV inverter)
	4 mH (1500 VA PV inverter)
Capacitance	0.25 μ F (500 VA PV inverter)
	1 μ F (1000 VA PV inverter)
	1.50 μ F (1500 VA PV inverter)
Damping resistor	20 Ω (500 VA PV inverter)
	30 Ω (1000 VA PV inverter)
	40 Ω (1500 VA PV inverter)
Grid side inductance	45.4 mH (500 VA PV inverter)
	23 mH (1000 VA PV inverter)
	9 mH (1500 VA PV inverter)

maximum power to the grid without creating a voltage rise. In the conventional approach, direct axis components of voltage and current are extracted using the PID controller, which is tuned by the trial and error method, and tuning is a manual process. So, by using the PID controller, mitigation of voltage rise is not possible.

A general ANN structure has three layers, such as input layer, hidden layer and output layer. Each layer has neurons linked to each other, which are followed by a weight factor. Levenberg–Marquardt Back Propagation (LMBP) checks for the actual value with reference value and if there is an error, the ANN network automatically adjusts the weight of the neurons so as to minimize the error. This error is propagated backward through the network.

In the proposed methodology, three-feed forward ANN controllers are used for obtaining the actual values of extracted components of current from PV inverter output current via dq transformation. After that, actual values are compared with reference values of quadrature and direct axis components using the ANN controller for generating the derived components of voltage by the proper machine-learning algorithm.

The Artificial Neural Network current controller always tries to minimize the difference between actual values and reference values of extracted components of current, as shown in Fig. 28.5. This leads to the derivation of the components of voltage for reference sine wave with the help of ANN-based controllers. This can be

Fig. 28.5 Proposed control algorithm

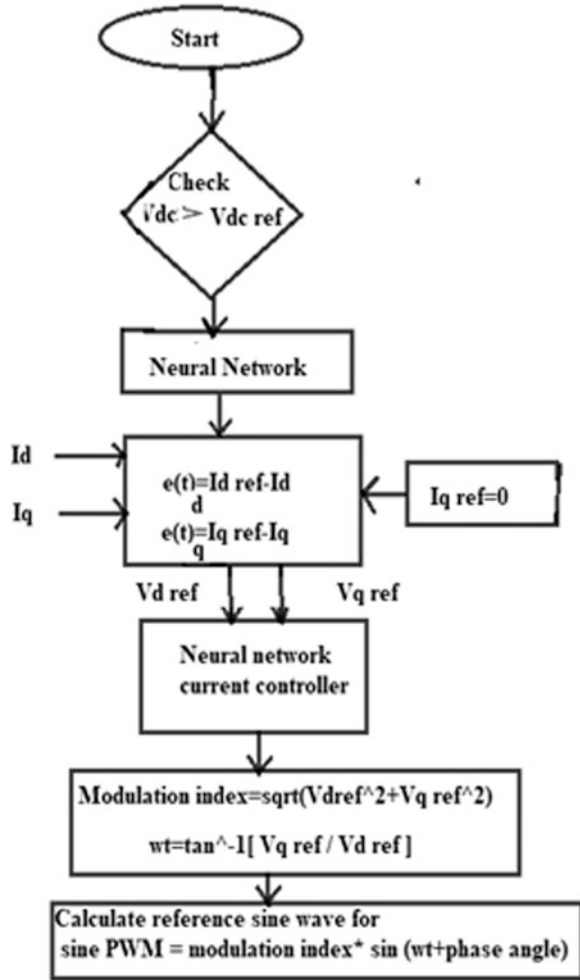


Table 28.2 ANN parameters

Testing Data	4
Training data	2
Hidden neuron	10
No of iterations	1000

used for PWM generation for inverter. The artificial neural network parameters used in this chapter are given in Table 28.2.

In this control scheme, a single-phase pulse width modulated voltage source inverter can be controlled using Levenberg–Marquardt Back Propagation (LMBP) ANN technique so as to mitigate the voltage rise issue in distribution networks connected with distributed PV systems, as shown in Fig. 28.6. In this approach,

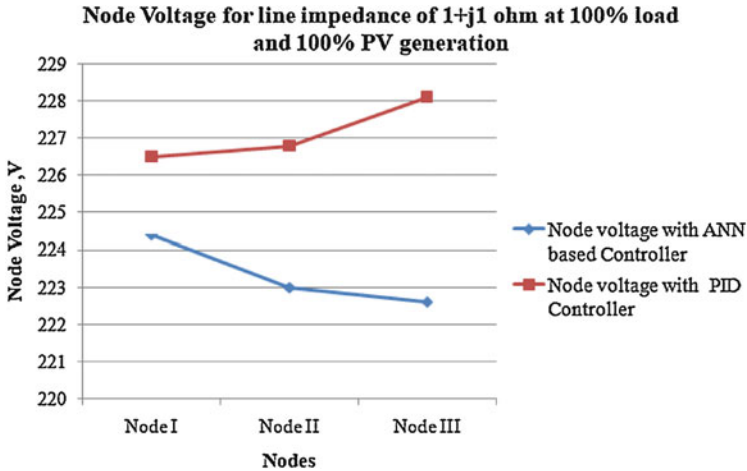


Fig. 28.7 Estimation of Node Voltages for 100% Load at 1000 W/m² irradiation with PID and ANN-based Controller

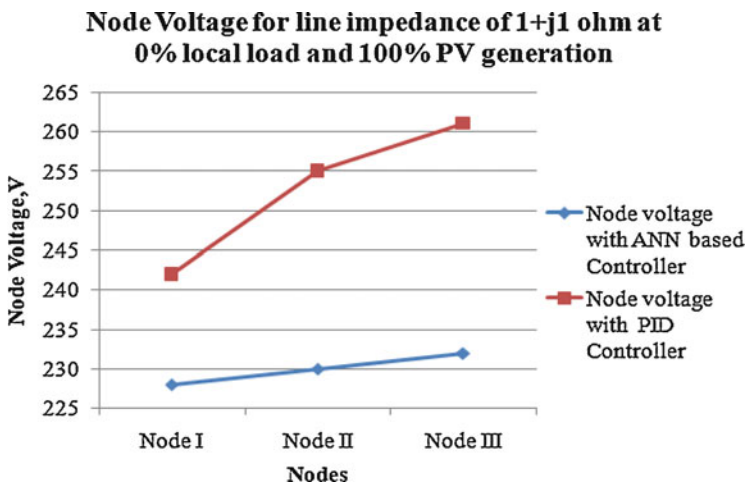


Fig. 28.8 Estimation of node voltages for 0% load at 1000 W/m² irradiation with PID and ANN-based controller

28.5 Conclusions

The effects of load variation changes in irradiation of solar PV system integrated with single-phase low-voltage distribution network are studied in this chapter. A typical distribution system with only three nodes having an X/R ratio of 1 is considered. Two types of control algorithm, mainly conventional PID and ANN based, are considered for pulse generation of a H-bridge converter.

From the analysis, it is observed that voltage rise occurred when the local load varies from full load to no load. From the simulation results, it is understood that ANN-based controllers should extract the maximum output power from solar PV systems while maintaining the node voltage at the rated voltage. In this chapter, the effect of line impedance, local demand on the voltage profile in distribution networks with distributed solar PV systems and the voltage-rise issue in the highly penetrated PV systems are studied and addressed. This chapter suggested a control strategy based on ANN algorithm for eliminating the voltage rise issue at the feeder line. When the penetration of PV systems is higher, chances for this voltage rise are more, and the ANN-based approach is proved to be a more efficient and effective technology to mitigate the voltage rise issue in a low-voltage distribution network.

References

1. Han CH, Ko BK, Song SY, Jan GS (2017) Application of PV Inverter on distribution system with high penetration of PV. *J Int Council Electr Eng* 7(1):263–268
2. Kumratug P (2012) Effect of R/X Ratio of the short transmission line on transient stability. *Am J Appl Sci* 3:365–367
3. Erns B, Engel B (2012) Grid integration of distributed PV-generation. In: Proceedings of the IEEE power energy Society general meeting, San Diego, CA, USA, July 2012, pp 1–7
4. Devidas AR, Korada P (2016) Studying the impact of AC-micro grid on the main grid and it's fault analysis. In: Proceedings of 2016 Biennial International Conference on Power and Energy Systems: Towards Sustainable Energy (PESTSE), January 2016
5. Haque MM, Wolfs P (2016) A review of high PV penetrations in LV distribution networks: Present status, impacts and mitigation measures. *Renew Sustain Energy Rev* 62:1195–1208
6. Masters CL (2002) Voltage rise: the big issue when connecting embedded generation to long 11 kV overhead lines. *Power Eng J* 16(1):5–12
7. Sargar DS, Bhattar CL (2015) LCLR filter design and modelling for harmonic mitigation in interconnected micro grid system. *Int J Res Eng Technol*, PISSN: 2321-730, April 2015
8. Sindhu MR, Nair MG, Sindhu S (2016) Photovoltaic based adaptive shunt hybrid filter for power quality enhancement. *IEEE International conference on power electronics, Drives and energy systems (PEDES)*, December 2016
9. Ari GK, Baghzouz Y (2011) Impact of high PV penetration on voltage regulation in electrical distribution systems. *International conference on clean electrical power*, June 2011, pp 744–748
10. Ghiani E, Pilo F (2015) Smart inverter operation in distribution networks with high penetration of photovoltaic systems. *J Modern Pow Syst Clean Energy*. <https://doi.org/10.1007/s40565-015-0165-4>
11. Zeraati M, Golshan MEH, Guerrero JM (2018) Distributed control of battery energy storage systems for voltage regulation in distribution networks with high PV penetration. *IEEE Trans Smart Grid* 9(4):3582–3593
12. Shinde SU, Sharmila M, Patil RS, Malkhede DV (2016) Performance comparison of PI & ANN based STATCOM for 132 KV transmission line. In: Proceedings of International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) June 2016
13. Tey LH, So PL, Chu YC (2003) Neural network-controlled unified power quality conditioner for system harmonics compensation. *IEEE/PES Transmission and Distribution Conference and Exhibition*, 19 February 2003
14. Mekhilef S, Mokhlis H, Karimi M (2016) Mitigating methods of mitigating power fluctuations of photo voltaic sources-a review. *Renew Sustain Energy Rev* 59:1170–1184

Chapter 29

Control of Buck Converter by Fuzzy Controller for Wind Energy: Battery System



Sheeba Babu and L. Ashok Kumar

Abstract The scope for wind farm as assessed using GIS platform is found to be more than 2000 GW. At present, India has installed wind farms of approximately 28,000 MW capacity and stands fourth largest in the world. So, the scope of improvement in the figures is immense. As a sustainable power source, wind source has turned out to be a promising one. Much research and development in wind energy conversion systems (WECS) has shown their excellent potential in remote areas located so far from power stations and distribution networks where it is uneconomical to install them.

As the output from renewable sources is highly variable, the battery system holds the key to stabilizing it before its integration with load. This chapter presents an introduction to fuzzy logic-controlled buck converters powered by wind energy conversion systems for constant voltage battery charging. It also presents the simulation analysis and comparison between PI controller and fuzzy controller operations of buck converters.

Keywords DC–DC Buck Converter · PI Controller · Fuzzy Logic Controller · Wind Energy Conversion System · Permanent Magnet DC Generator

Abbreviations

FLC	Fuzzy logic controller
GIS	Geographic information system
MF	Membership function

S. Babu
PSG College of Technology, Coimbatore, Tamil Nadu, India

L. Ashok Kumar (✉)
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

MPPT	Maximum power point tracking
PI	Proportional integral
PMDC	Permanent magnet DC generator
WECS	Wind energy conversion system

29.1 Introduction

Energy serves as the key factor for development of any country. Energy demand for the time being is mostly met by fossil fuels. At the present rate of consumption, the fossil fuels will get exhausted within a few decades. Moreover, carbon emission adversely affects the climatic conditions [1]. So, it is the need of the hour to substitute fossil fuels by sustainable sources of energy to support the load demand.

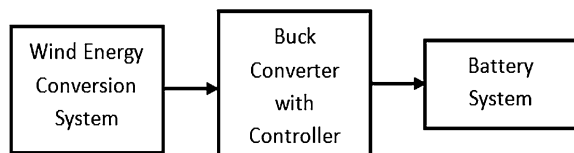
Since digital technology upgrades the lifestyle of humans and makes their life comfortable and enjoyable, it turns out to be most accepted among all [2–5]. These days, people cannot imagine a world devoid of electronic appliances like laptops, mobile phones and stereo systems. Along with these technological advancements comes the need of control system with intelligence to cope up with the ever-increasing demand for customer satisfaction [6–11].

Apart from the Introduction, this chapter is spread over five sections. Section 29.2 presents an overview of the proposed system. Section 29.3 provides an introduction to DC–DC converters, classical controllers and fuzzy structure with an insight into the proposed system. In Sect. 29.4, an overview of fuzzy logic controllers is given. Section 29.5 explains the simulation block diagram. Simulation results are presented in Sect. 29.6. Section 29.7 provides conclusions and further work suggestions for controlling the voltage of wind energy battery systems by using DC–DC converters.

29.2 Overview of the Proposed System

Figure 29.1 gives the block diagram of the proposed system. Mainly at far-off locations where it is too expensive to establish supply networks, standalone systems are preferred [3]. A realistic approach for standalone power generation is to use a battery system along with a variable-speed permanent-magnet DC generator (PMDC)-driven wind turbine to create an autonomous system. PMDCs are used widely in standalone WECS because of their advantages: higher reliability, less maintenance and more effectiveness. Besides, it is suitable for variable-speed operations, which provide 10–15% higher energy output, lower mechanical stress and less power

Fig. 29.1 Block diagram of proposed system



fluctuation compared with constant-speed operations. A system of variable-speed PMDC wind turbines has more flexibility because it can adapt to wind variations [4–6]. In this chapter, a direct-searching MPPT controller is designed for variable-speed PMDC-driven wind turbines. For the reason that conventional PI controllers generally do not work well for non-linear systems, high-order systems and complex systems that have no precise mathematical models [7], fuzzy logic is chosen to overcome these difficulties due to its flexibility and facility to apply in real-time control [8]. To provide stability to the autonomous system, the battery system is recommended in a standalone mode. The wind generator gives power to meet the load and due to the intermittency of source and variable-load demand, the battery system bridges the gap. If power is in excess of demand, the battery is charged and if power is deficit, it is discharged. With this proposed method, battery charging at constant voltage is ensured irrespective of input voltage variation in generated voltage.

29.3 Buck Converter

DC–DC converters with digital control methods became popular due to their mass production, low power consumption, superior resistance to ambient variations, like temperature, seamless interface to complex control strategy. Their requisitions include dense electronic components, such as laptops, mobile phones and others [12–15]. Diverse methods have been used to meet the necessity of output voltage, and simultaneously it is also essential to obtain further resolution to raise accuracy. As the application of various devices is different, the power supply, current and voltage rating would be diverse as per their requirement [14].

The loss of power in switched regulators is low due to the removal of the resistor. This factor renders the switching regulator more competent as compared with a linear regulator. Also, the output voltage of this regulator does not depend on the load. There is a variety of switching regulators but our concentration is limited to the buck regulator. The schematic diagram of the buck converter is shown in Fig. 29.2.

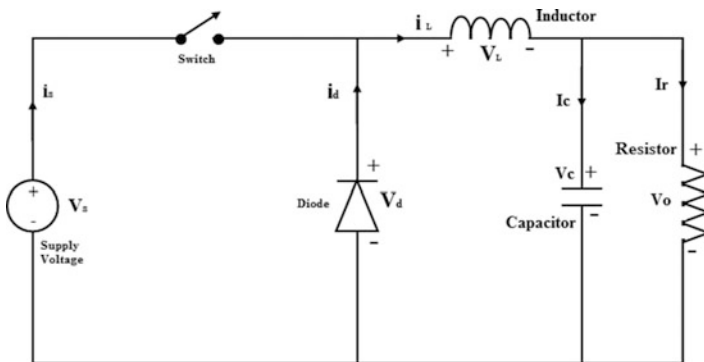


Fig. 29.2 Schematic diagram of buck converter

As this converter uses two power switches operated synchronously, that is, one after another, it is called a synchronous buck regulator.

The buck converter is a type of step-down chopper where the output voltage is less than the input voltage. Here the regulation is done using simple switches. These switches go ON and OFF at a fixed rate, called switching frequency, to keep the output at the desired level [4].

The output of the DC–DC buck converter is given as

$$V_o = D \times V_{in} \quad (29.1)$$

Where, D is the duty ratio, which is defined by the ratio of the on time of the switch (T_{ON}) to the operating time period (T).

29.4 Fuzzy Logic Controller (FLC)

The PI controller has been used in many industrial applications due to its ability to operate most practical industrial processes [15]. It consists of a proportional element and an integral element. The components are connected in parallel. K_p and K_i are the gains of the P and I elements, respectively. The error signal is considered the input of a PI controller. As the plant or system to be controlled is getting complex by the day, the mathematical modelling of the system is getting complicated. Hence the modern control system based on heuristic approach is gaining importance nowadays.

The use of FLC has also been very popular in control systems. The reason for increasing the use of FLC is because of its ability to use a linguistic form instead of mathematical form to manipulate knowledge. FLCs also have other advantages such as low cost and simplicity of control, which makes them more used than other classical controllers in many control systems. It is the most important control method for nonlinear systems [1].

There are three principal elements to a fuzzy logic controller, as shown in Fig. 29.3. These elements are the fuzzification module (fuzzifier), the rule base and the defuzzification module (defuzzifier). The fuzzifier converts the crisp values

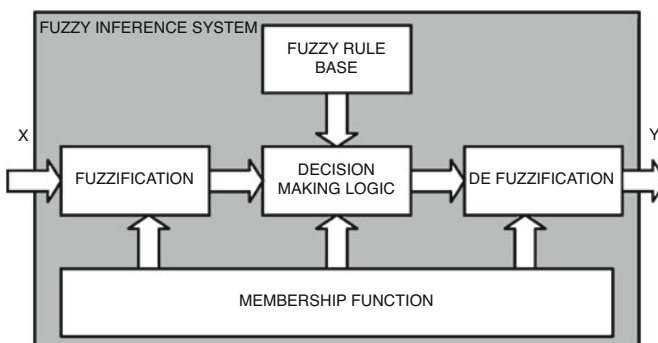


Fig. 29.3 Principle elements of Fuzzy Logic Controller

of the input into fuzzy values to send to the rule base. The rule base is expressed as a set of if-then rules based on predetermined expert knowledge. The defuzzifier converts the output values of the rule base to the crisp values [1, 11].

In Fig. 29.3, X is the input data. Y is the output data. Membership function is the type of function that is used in fuzzy control, such as triangular MF and Gaussian MF. There are various types of membership functions used in fuzzy set theory, including trapezoidal MFs, Gaussian MFs and generalized bell MFs.

In the fuzzy controller for the considered power converter, there are two input variables: the error eu and the change of error deu [1-2-5]. The three triangular membership functions are chosen for two input variables and an output variable, as shown in Fig. 29.4.

In our case, the input membership functions are triangular with five positions. The input parameters are error eu ($V_{ref}-V_{mes}$) and the change of error deu ($eu(k)-eu(k-1)$). The output membership function is also triangular with seven positions. The output parameter is the change to bring to the duty cycle y ($a(k) = a(k-1) + da(k)$). Inference rules are determined depending on the behaviour of the converter and are shown in Table 29.1. Defuzzification of output parameter (namely change of duty

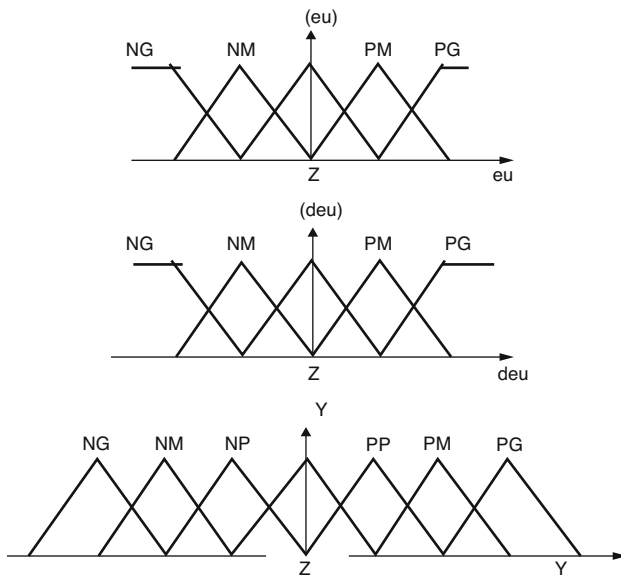


Fig. 29.4 Membership functions for input and output

Table 29.1 Inference rules

eu/ deu	NG	NM	Z	PM	PG
NG	NG	NG	NG	NG	NM
NM	NG	NM	NP	Z	PP
Z	NM	NP	Z	PP	PM
PM	NP	Z	PP	PM	PG
PG	PM	PG	PG	PG	PG

cycle) is done with centroid method. The fuzzy controller for the buck converter is performed in the MATLAB–SIMULINK environment. At first, we set inference rules and membership functions for the three variables error (eu), change of error (deu) and the change of the duty cycle in the MATLAB fuzzy toolbox, and simulate the system to have a fuzzy inference matrix (FISmatrix) that we can use in the SIMULINK environment with the block called fuzzy logic controller.

29.5 Simulation Block Diagram

The simulation study and analysis are conducted in three stages. Figure 29.5 shows the block diagram of the buck converter controlled by the PI controller. The variable input is generated with controlled voltage source block. A stair generator block with rise limiter and saturator is used to provide input to control the signal of the controlled voltage source block. The stair generator changes its output abruptly at a set time to a set value. But in the real world, the wind speed does not change abruptly—it changes at a faster or slower rate—an average rate is included by the rate limiter block. Instead of battery, a resistive load is used to check the output voltage of the buck converter. The output voltage is compared with the set voltage used for charging the battery, and the error is used by the PI controller to control the on and off time of the power switch.

Figure 29.6 shows the block diagram of the buck converter controlled by FLC. The output voltage is compared with the set voltage used for charging the battery. The operation of FLC is explained in Sect. 29.4. The error and change in error are the

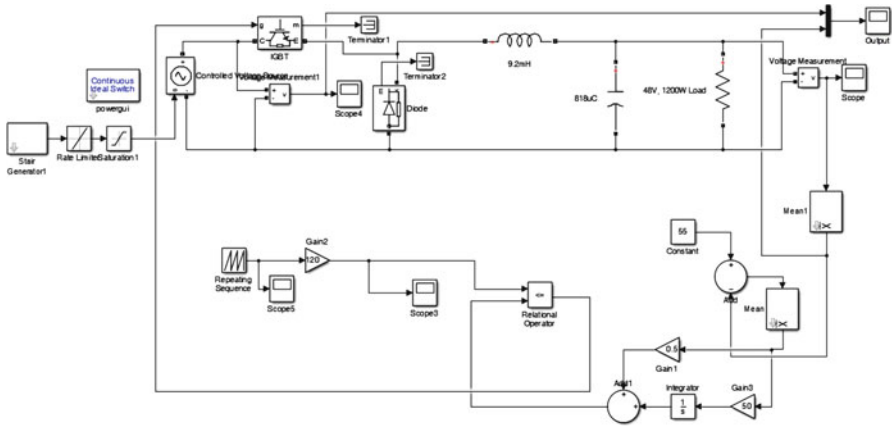


Fig. 29.5 Buck converter controlled by PI controller

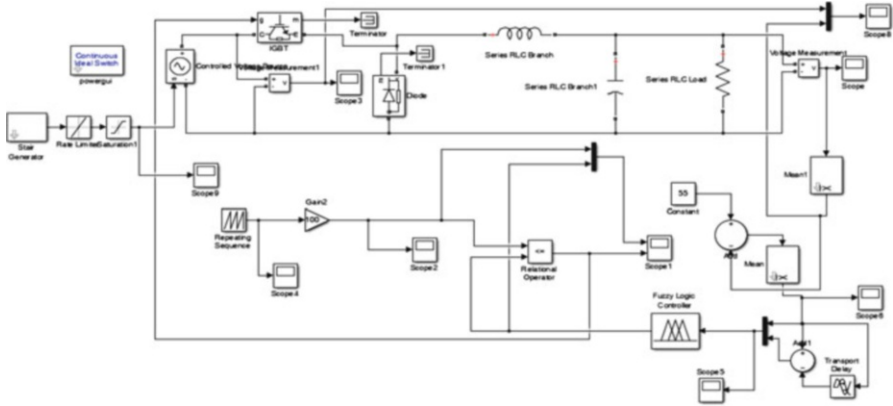


Fig. 29.6 Buck converter controlled by FLC

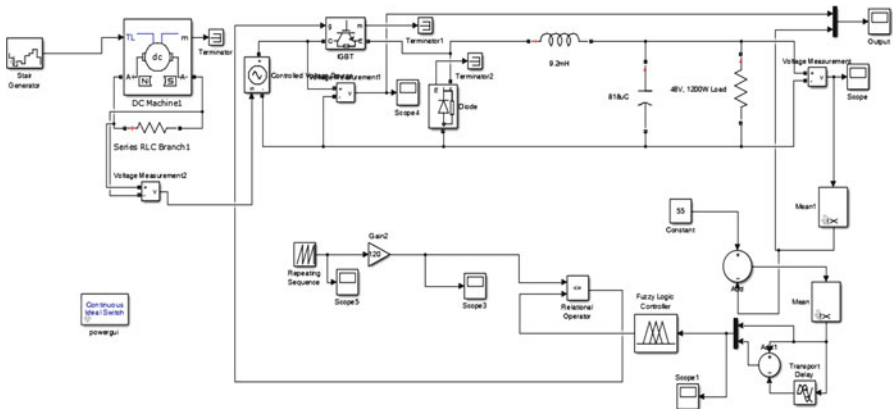


Fig. 29.7 Buck converter controlled by FLC with PMDC as wind energy converter

two inputs given to the FLC block to generate the reference signal to control the on and off time of the power switch.

Figure 29.7 shows the block diagram of the buck converter controlled by FLC with PMDC as wind energy converter. PMDCs are used widely in standalone WECS because of their advantages: higher reliability, less maintenance and more effectiveness. The output voltage of PMDCs is used as the control signal for controlled voltage source. As PMDC is a rotating machine, it has got rotational inertia to take

care of abrupt changes in input. The PMDC system is so designed to produce rated output at rated torque.

The response is discussed in the section to follow, which highlights the effectiveness of the controller.

29.6 Simulation Results

This section gives the behaviour of the converter for the above-mentioned block diagram. Figure 29.8 shows the input and output of the buck converter with a PI controller for regulating the output for charging the battery. For constant-voltage battery charging, the permissible variation in output voltage is $\pm 10\%$. As can be seen from the output plot, the PI controller has got a very good transient response. But at some instances, it crosses the permissible band of the 10% variation from the set value.

Figure 29.9 shows the input and output of the buck converter with FLC for regulating the output for charging the battery. As can be seen from the output plot,

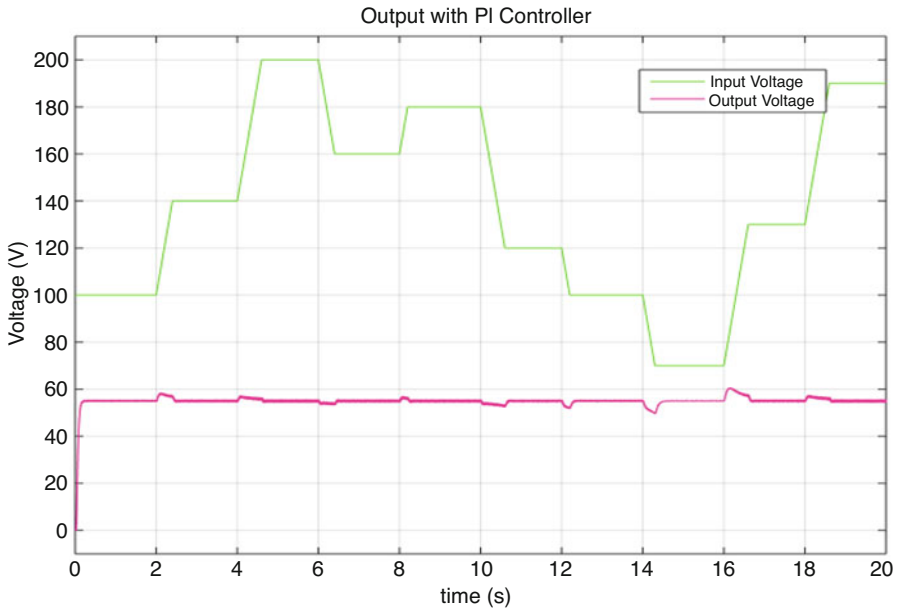


Fig. 29.8 Buck converter input and output with PI controller

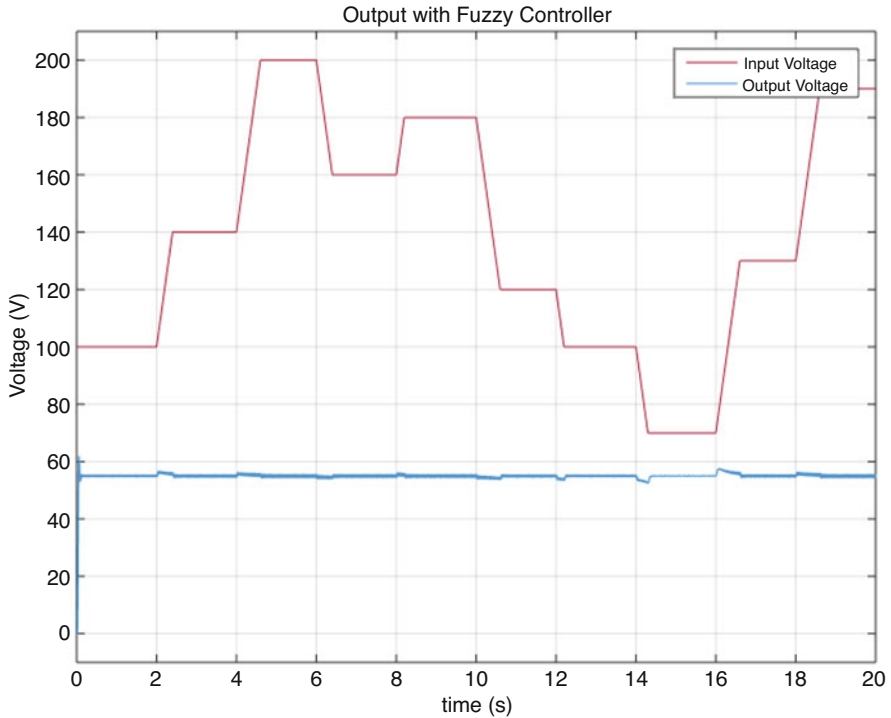


Fig. 29.9 Buck converter input and output with FLC

FLC has got a transient overshoot. But a very good regulation within the permissible band of variation.

Figure 29.10 shows the comparison in performance of the PI controller and FLC, as is clear from the output voltage waveform; apart from the transient voltage overshoot at the initial point, FLC shows much better performance to PI controller.

Figure 29.11 shows the buck converter input and output voltage waveform with PMDC, as PMDC has rotational inertia the rate of rise and fall of which is much lower and hence the output is almost a constant with a very small overshoot but an appreciable rise time due to inertia of rotating machine.

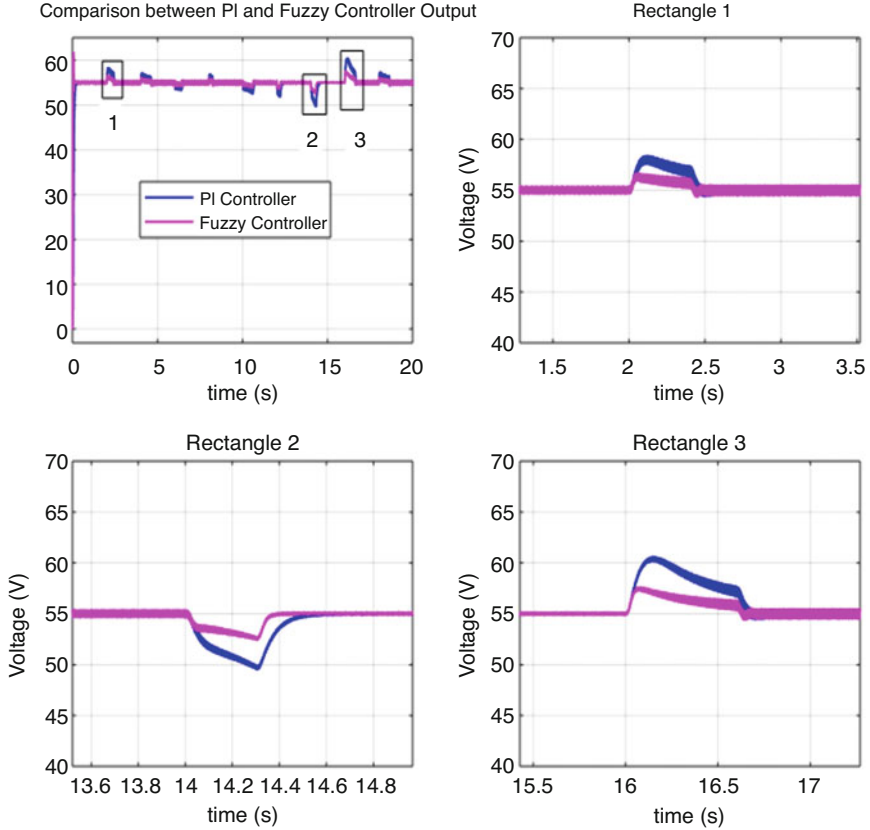


Fig. 29.10 Comparison of FLC and PI Controller output for the same input voltage

29.7 Conclusion

This chapter presents an analysis of a buck converter by simulation. A closed-loop model was implemented in MATLAB–SIMULINK toolbox SIMPOWERSYS. The same power circuit was controlled by two different controllers, namely PI controller and FLC. The results were compared and FLC was found to have better performance. The study proves the robustness of the fuzzy controller. It gives better stabilization without any need for mathematical model or design process.

The hardware implementation of the same can be done for validating the results.

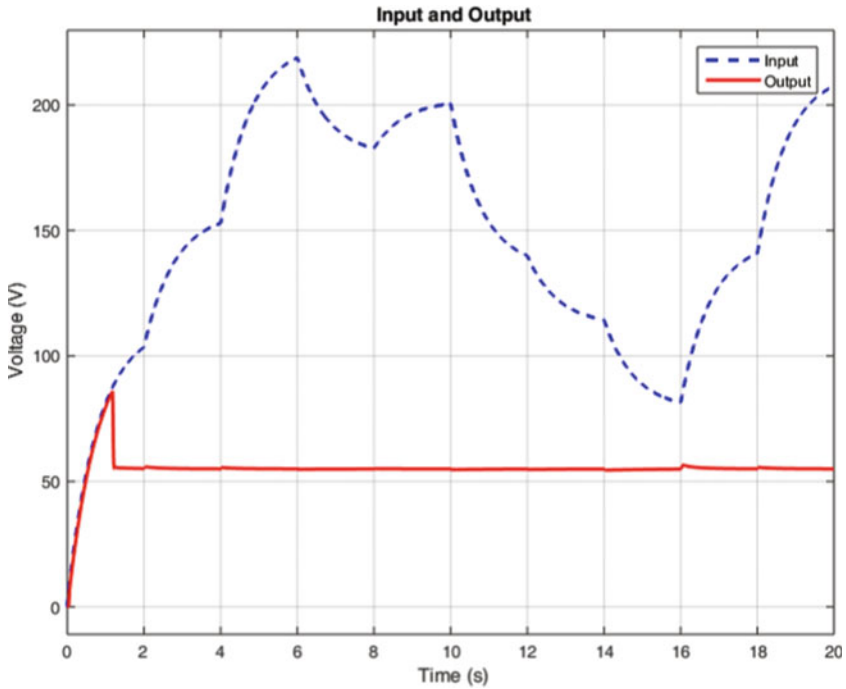


Fig. 29.11 FLC controlled buck converter input and output with PMDC

References

1. Mattavelli P, Rossetto L, Spiazzi G, Tenti P (1997) General purpose fuzzy controller for DC-DC converters. *IEEE Trans Power Electr* 12(1):79–86
2. Kovačević H, Stojanović Ž (2016) Buck converter controlled by Arduino Uno. *MIPRO 2016*, May 30–June 3, 2016, Opatija, Croatia, pp 1638–1642
3. Singh N, Dr Singh B (2016) Design and modeling of wind energy conversion system based on PMSG using MPPT technique. *Int J Sci Res Eng Technol (IJSRET)* 5(2):96–100
4. Kumar A, Paliwal NK, Singh AK, Kumar P, Sehgal S, Singh NK, Singh RK (2017) Optimal energy management of hybrid power system with two different battery models. *2017 International Electrical Engineering Congress (iEECON)*
5. Mazouz N, Midoun A (2011) Control of a DC/DC converter by fuzzy controller for a solar pumping system. *Electr Power Energy Syst* 33:1623–1630
6. Elmas C, Deperlioglu O, Sayan HH (2009) Adaptive fuzzy logic controller for DC–DC converters. *Expert Syst Appl* 36:1540–1548
7. Kumar LA, Babu S (2016) Control strategy for inverter in a micro-grid under unbalanced grid condition. *2016 IEEE conference on technologies for sustainability (SusTech)*
8. Dr Jos BM, Abhijith S, Venugopal A, Roy B, Dhanesh R (2014) Fuzzy logic controlled PV powered Buck converter with MPPT. *Int J Adv Res Electr Electr Instrum Eng* 3(5):9370–9377

9. Smyej M, Cheriti A (1999) Fuzzy logic controller for a DC to DC converter. In: Proceedings of the 1999 IEEE Canadian conference on electrical and computer engineering Shaw conference center, Edmonton, Alberta, Canada May 9–12 1999, pp 1020–1023
10. Uma Siva A, Jyothi DS, Gopal P, Ramu G (2013) Bi-directional dc-dc converter drive with PI and fuzzy logic controller. *Int J Adv Res Electr Electr Instrum Eng* 2(11):5435–5441
11. Sahin ME, Okumus Hİ (2011) Fuzzy logic controlled buck-boost DC-DC converter for solar energy-battery system. IEEE conference paper, June 2011, pp 394–397
12. Corcau JI, Dinca L (2016) Fuzzy logic control for a DC to DC Buck converter. *Scientific research and education in the air force-AFASES 2016*, pp 233–238
13. Sasikumar S, Krishnamoorthi K, Sivakumar R (January 2017) A bidirectional converter using fuzzy controller. *Int J Res Appl Sci Eng Technol (IJRASET)* 5(1):429–436
14. Ugale CP, Dr VVDixit (2017) Buck-boost converter using fuzzy logic for low voltage solar energy harvesting application. 11th International Conference on Intelligent Systems and Control (ISCO), pp 413–417
15. Renwal D, Kumar M (2015) Hybrid PI-fuzzy logic controller based DC-DC converter. 2015 International Conference on Green Computing and Internet of Things (ICGCIoT) IEEE, pp 753–757

Chapter 30

A Survey on Secure Beamforming in MIMO-NOMA-Based Cognitive Radio Network



Thulasimani Lakshmanan and Hyils Sharon Magdalene Antony

Abstract Cognitive radio network (CRN) and non-orthogonal multiple-access (NOMA) make a significant contribution to fifth-generation (5G) wireless communication systems. But securing multiple-input–multiple-output (MIMO) NOMA beamforming is yet an exclusive way. This chapter reveals the latest survey of the security of MIMO NOMA in 5G engineering, which includes securing by cascaded transmitting downlink zero-forcing-beamforming (ZFBF) technique, general power allocation scheme, applying the NOMA protocol in large-scale networks and applying new bi-directional ZFBF mechanism, cooperative NOMA in both amplify-and-forward (AF) and decode-and-forward (DF) protocols. Also an efficient majorization–minimization (MM) method-based semi-closed form secrecy rate optimization algorithm are reviewed in this chapter.

Keywords Cognitive radio network (CRN) · Non-orthogonal multiple-access (NOMA) · zero-forcing-beamforming (ZFBF) · AF and DF protocols · Majorization-minimization (MM) method

Abbreviations

5G	fifth generation
AF	amplify and forward
BSs	base stations
CoBF	coordinated beamforming
CRN	Cognitive radio network

T. Lakshmanan (✉) · H. S. M. Antony
PSG College of Technology, Coimbatore, Tamil Nadu, India
e-mail: ltm.ece@psgtech.ac.in

CSI	channel state information
D NOMA	Dynamic non-orthogonal multiple-access
DF	decode and forward
MIMO	multiple-input–multiple-output
MISO	multiple input single output
MM CVX	majorization minimization convex optimization
MM SCF	majorization minimization semi-closed form secrecy rate optimization
MM	majorization–minimization
NOMA	non-orthogonal multiple-access
OMA	orthogonal multiple-access
PL	physical layer
PU	primary user
SIC	successive interference cancellation
SNR	signal-to-noise ratio
SOP	secrecy outage probability
SU	secondary user
ZF NOMA	zero forcing non orthogonal multiple-access
ZFBF	zero forcing beamforming

30.1 Introduction

The NOMA technique leads to high throughput and wide connectivity with very low latency to reprocess the resources in both temporal and spectral varieties [1, 2]. It allocates the same frequency to different users within the same cell. Then, NOMA is integrated with multiple-input–multiple-output (MIMO) to attain high spectral efficiency in a multi-cell framework. The serious interference makes the system open to security threat. In the general MIMO-NOMA Scheme [3, 4], signal alignment and the receiving beamformer are used to overcome the interference and increase the total throughput. In the beamforming technique [5], the transmitting and the receiving techniques are taken to lessen inter- and intra-cell interferences. The receiving beamforming cannot secure the data as eavesdropper(s) may tap the channel merely by ignoring the receiving beamformers. The MIMO-NOMA system model [6] has artificial noise to secure the transmission in a single cell. The approach for convex optimization problem was found in [7]. In [8], the relay-based downlink NOMA network security was considered. Thus, the received beamformer technique in [3]–[7] does not guarantee system security in a multi-cell framework. In this chapter, six papers are reviewed for better security performance of MIMO-NOMA beamforming in CR Network.

30.2 Study of Securing MIMO NOMA in 5G Technology

30.2.1 Defence for MIMO-NOMA-Based CRN

A downlink cascaded transmitting ZFBF technique is suggested in this chapter in a 2-cell MIMO, NOMA-based CRN. By applying ZFBF, two base stations (BSs) design their transmitting ZFBF vectors to elevate physical-layer (PL) security. In Fig. 30.1 [9], this method is compared with the coordinated beamforming (CoBF) method [5]. The CoBF system hires both transmitting and receiving beamformers. The channel matrix is lined up by the transmitting beamformer method and is annulled by the receiving beamformer. For the CoBF technique, the secrecy rate does not rise with SNR [9].

In Fig. 30.2 [9], the secrecy outage probability of secondary user (SU) \gg (is greater than) primary user (PU). If the transmit signal-to-noise ratio (SNR) is higher, the secrecy outage probability is zero, which follows the asymptotic results.

Table 30.1 shows the various contributions, channel state information (CSI) and the system model used in various research papers to improve secrecy capacity.

30.2.2 A General Power Allocation Scheme for NOMA

The chapter proposes [10] a new dynamic power allocation system to uplink and downlink NOMA assumptions in two users. The outage probability rating and the

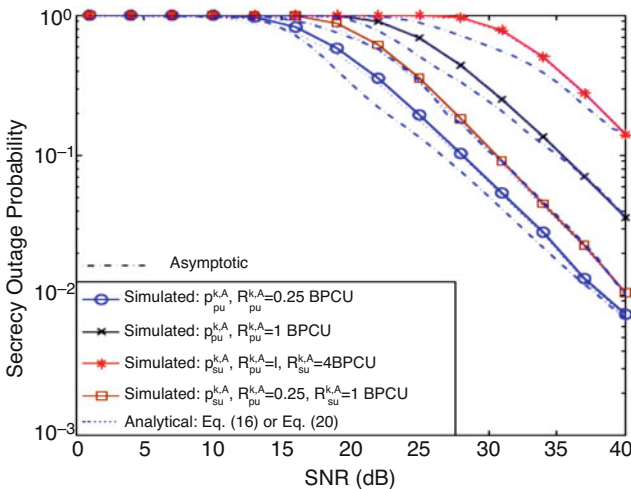


Fig. 30.1 Performance comparison of the scheme with CoBF

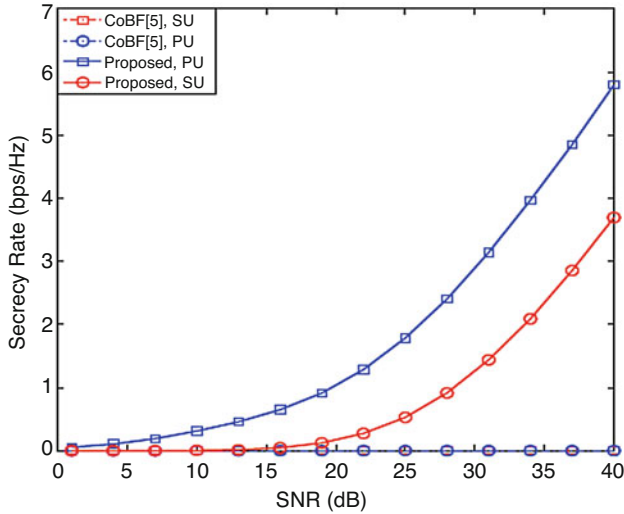


Fig. 30.2 The asymptotic outage performance of PU and SU

mean rate of the system with estimations of high SNR are detected. In Fig. 30.3, the D-NOMA scheme outperforms OMA due to the fact that D-NOMA achieves better spectral efficiency [10]. D-NOMA avoids the user with poor channel conditions, which is smaller than that of orthogonal multiple-access (OMA) [10].

30.2.3 *Enhancement of NOMA's Physical Layer Security in Large-Scale Networks*

The chapter [6] tells about the physical-layer security of NOMA in large-scale networks. Single- and multiple-antenna transmissions are regarded on the BS to communicate with the allotted NOMA users. In the single-antenna assumption, m NOMA users cast in a limited disc or user partition with the tone-arrangements of their channel. For instance, the m -th NOMA user indicates the m channel-quality arrangements. In this event, the m -th and n -th users are paired for transmission among the similar resource slot. In the multiple-antenna assumption, beamforming is call down at the BS to generate artificial noise [6]. The complexity of channel ordering [6] of multiple-input–single-output (MISO) channels in NOMA is reduced by dividing the round cell (in Fig. 30.4) into an internal and an external disk. Then one user is chosen from the internal disc and another from the external disk. Using a NOMA protocol, the two users are paired for transmission. In Fig. 30.5, the disc D is isolated into two parts, viz., $D1$ and $D2$. The first one creates channel quality differences between the paired users. The second one reduces the channel ordering complexity in this MISO-NOMA system.

Table 30.1 Survey of various methods in securing MIMO-NOMA-based CRN

Paper	System model	Main channel	CSI	Objective	Criteria	Main contribution
N. Nandan et al. [9]	MIMO-NOMA-based CRN, multi-cell, multi-cluster environment, and multi-eavesdropper, in Rayleigh flat fading channel	Gaussian	Perfect CSI of the wanted user, and statistical CSI of the secret listener or eavesdropper	The total secrecy rate is increased in the system	Secrecy outage probability	A downlink cascaded ZFBF technique to maximize the secrecy outage probability in a cluster
Z. Yang et al. [10]	In uplink and downlink NOMA, in multi-user, and in multiple single-antenna eavesdroppers	Rayleigh fading channel gain	Global perfect CSI no CSI at the eavesdropper	Dynamic NOMA (D-NOMA) realizes a more balanced user and is compared to the fixed-NOMA and CR-NOMA	Outage probability and, average rate of perfect users	Analyse the power allocation scheme of NOMA with dynamic power allocation, D NOMA
Y. Liu et al. [6]	NOMA in large-scale networks	MIMO wiretap channels, Gaussian as 'eves' channel	Perfect CSI of users, and no CSI at the eavesdropper	Enquire the secrecy performance of large-scale NOMA networks in a single-antenna aided and a multiple-antenna assisted at the BS	Secrecy outage probability (SOP)	The secrecy performance of applying NOMA protocol in large-scale networks was examined
N. Nandan et al. [11]	MIMO-CRN	Rayleigh flat-fading channels	Perfect CSI of the desired users and statistical CSI at the eavesdropper	Overall secrecy capacity of the network is improved	Secrecy capacity, ergodic secrecy capacity, and, secrecy outage probability	A bidirectional ZFBF mechanism at the transmitter analyse the secrecy capacity and secrecy outage of PU, SU and the underlay MIMO-CRN
J. Chen et al. [8]	Cooperative NOMA system	Rayleigh fading channel	Global perfect CSI	Secrecy outage probability (SOP) tends to constant at very high signal-to-noise ratio (SNR). For	SOP, strictly positive secrecy capacity	The secrecy performance of cooperative NOMA in AF and DF protocols are

(continued)

Table 30.1 (continued)

Paper	System model	Main channel	CSI	Objective	Criteria	Main contribution
M. Jiang et al. [7]	Cellular downlink MIMO NOMA	Gaussian	No CSI at the transmitter	at high SNR, the secrecy execution of AF and DF NOMA are same New NOMA scheme outperforms ZFBF NOMA scheme and a conventional OMA scheme	Achievable secrecy rate	independent for channel conditions between the relay and the poor users An efficient MM-based semi-closed secrecy rate optimization algorithm for secure network transmission

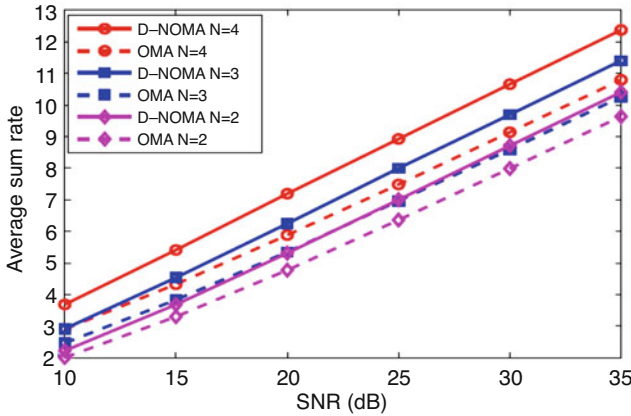


Fig. 30.3 Average sum of downlink D-NOMA systems compare to OMA in different users, $M = 4$

Fig. 30.4 Network framework for secure NOMA transmission in a single-antenna assumption, where r_p , R_D is the radius of the Eve-exclusion area and NOMA user zone, and ∞ is an Eves infinite two-dimensional plane, respectively

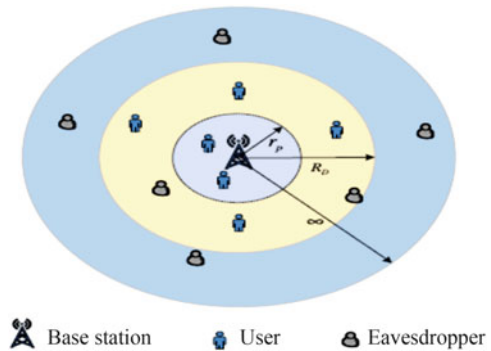
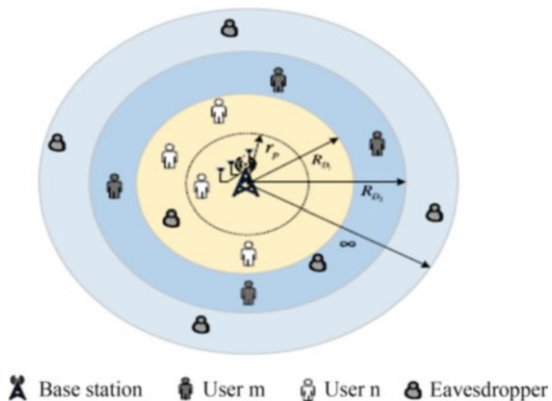


Fig. 30.5 Network framework for secure NOMA transmission using an artificial noise with a multiple antenna has been considered and r_p is the radius for Eve exclusion zone. R_{D1} is the NOMA user zone for user n , and R_{D2} is the NOMA user zone for user m , and ∞ is an eavesdroppers infinite two dimensional plane, respectively



30.2.4 *Secrecy Capability Maximization*

This chapter technique [11] shows the optimal PL security for both PU and SU in the underlay MIMO-CRN by the use of bidirectional ZFBF techniques. This enables the PU to be communicated with the SUs by a relay node. In the first phase, a transmitting beamformer at the PU/SU transmitters and a receiving beamformer at the relay node are assumed to separate PU and SU information. In the second phase, a bidirectional beamformer is employed to remove artificial noise. The total secrecy capacity is maximized by formulating a generalized non-linear optimized problem. The method [11] detects the presence of imprecise CSI between the intended transmitter and the eavesdropper. The underlay MIMO-CRN in this chapter consists of a primary transceiver, a secondary transceiver, a relay and an eavesdropper.

30.2.5 *Physical Layer Security of Cooperative NOMA Systems [8]*

For cooperative NOMA system security, both the amplify-and-forward (AF) and decode-and-forward (DF) protocols are conceived. The NOMA users are affected by the secret listener or an eavesdropper. AF and DF attain similar secrecy rate for greater SNR as shown in their results. An asymptotic result indicates that the SOP inclines to a fixed point have greater signal-to-noise ratio (SNR).

30.2.6 *Secure Beamforming in MIMO-NOMA [7]*

A cellular downlink MIMO-NOMA secure transmission network consists of BS, a central user and a cell-edge user. The central user is an entrusted user and the cell-edge user is a potential eavesdropper. The majorization–minimization (MM) method [7] was applied to get a sequence of valid surrogate functions in each of the iterations. This is called MM-semi-closed form secrecy rate optimization (MM-SCF) algorithm.

Assume the channel response [7] $H_i, i \in \{1, 2\}$, is modelled as $H_i = d_i^{-2} \mathbf{H}_i$ where d_1 and d_2 denote the distances from the BS to the cell-edge and central users, and the entries in \mathbf{H}_i are identically complex random variables with zero mean and unit variance. Assume $d_1 = 200$ m and $d_2 = 100$ m. The power of additive Gaussian noises is $\sigma^2 = -70$ dBm. If not specified, assume the BS with $M = 4$ antennas. The cell-edge and central users are equipped with $N_1 = N_2 = 2$ antennas. In Fig. 30.6, achievable secrecy rate of the central user R_s versus achievable rate requirement of the cell-edge user r where maximum transmit power at the BS is $P_{\max} = 25$ dBm. Compare MM, SCF with three schemes, that is, MM, convex optimization (CVX)

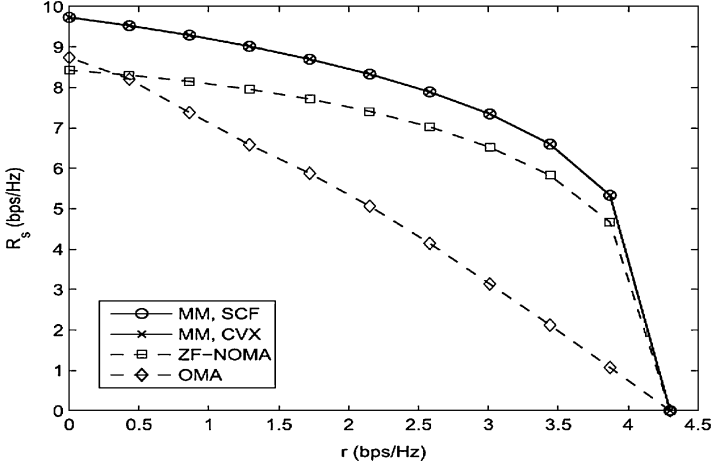


Fig. 30.6 Achievable secrecy rate of the central user, R_s versus achievable rate of the cell-edge user, r , where $P_{max} = 25$ dBm

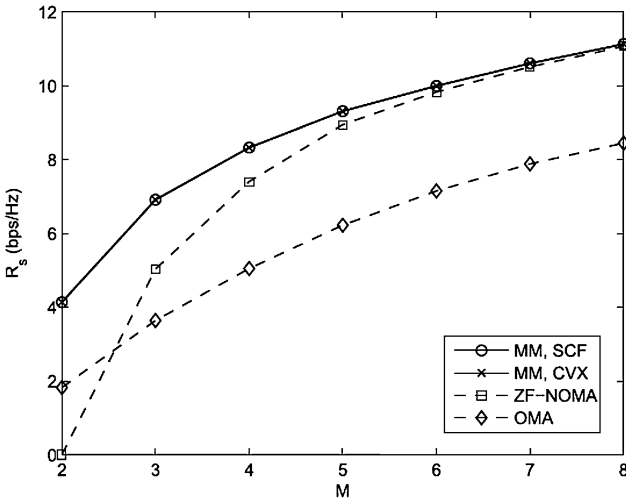


Fig. 30.7 Achievable secrecy rate of the central user R_{s2} , versus antenna M , where $P_{max} = 25$ dBm and $r = 2.1$ bps/Hz

method, ZF-NOMA and OMA Schemes. ZF-NOMA scheme refers to the NOMA scheme using ZF beamforming, that is, $Q_2H_1 = 0$ [7].

In Fig. 30.7, the performance gap of the proposed scheme and ‘ZF-NOMA’ is reduced with the increase of antenna M . The effect of ZF beamforming $Q_2H_1 = 0$ on the performance is reduced [7].

30.3 Future Guidance to Secure NOMA

The interference-alignment-transmitting ZFBF approach is proposed for the downlink MIMO-NOMA-based CRNs in which two BSs transmit ZFBF vectors to improve the PL security for both the cells. In future, multiple cells are presented, and each cell has a BS with a certain number of clusters, and a number of PUs and SUs. Each user has many number of antennas. The reduction in interference between the user and the other user introduces the ZFBF technique to improve system security. Using the mean square error scheme, one can maximize the secrecy probability and throughput of the system. The proposed increases the overall secrecy of the network. Another future direction [10] is to combine this novel D-NOMA power allocation scheme with clustering algorithms in a large-scale network. Assuming perfect successive interference cancellation (SIC) procedure may guide the overrate execution of the networks conceived, overall secrecy capacity of the network can be greatly improved.

30.4 Conclusion

This chapter provides efficient security for 5G MIMO-NOMA technology to enhance the secrecy rate and user capabilities. Beamforming security and general power allocation techniques develop an integrated security solution together efficiently safeguards the confidential and privacy communication data in MIMO-NOMA-based CR. The research result of review of the various papers is much useful for the researchers to get a better opportunity for securing MIMO-NOMA-based CR in this era of wireless networks.

References

1. Islam SMR, Zeng M, Dobre OA, Kwak KS (2018) Resource allocation for downlink noma systems: key techniques and open issues. *IEEE Wireless Commun* 25(2):40–47
2. Zeng M, Yadav A, Dobre OA, Tsiropoulos GI, Poor HV (2017) Capacity comparison between MIMO-NOMA and MIMO-OMA with multiple users in a cluster. *IEEE J Sel Areas Commun* 35(10):2413–2424
3. Ding Z, Schober R, Poor HV (2016) A general MIMO framework for NOMA downlink and uplink transmission based on signal alignment. *IEEE Trans Wireless Commun* 15 (6):4438–4454
4. Ding Z, Adachi F, Poor HV (2016) The application of MIMO to nonorthogonal multiple access. *IEEE Trans Wireless Commun* 15(1):537–552
5. Shin W, Vaezi M, Lee B, Love DJ, Lee J, Poor HV (2017) Coordinated beamforming for multi-cell MIMO-NOMA. *IEEE Commun Lett* 21(1):84–87
6. Liu Y, Qin Z, Elkashlan M, Gao Y, Hanzo L (2017) Enhancing the physical layer security of non-orthogonal multiple access in large-scale networks. *IEEE Trans Wirel Commun* 16 (3):1656–1672

7. Jiang M, Li Y, Zhang Q, Li Q, Qin J (2017) Secure beamforming in downlink MIMO nonorthogonal multiple access networks. *IEEE Sig Process Lett* 24(12):1852–1856
8. Chen J, Yang L, Alouini MS (2018) Physical layer security for cooperative NOMA systems. *IEEE Trans Veh Technol* 67(5):4645–4649
9. Nibedita N, Sudhan M, Hsiao-Chun W (2018) Secure beamforming for MIMO-NOMA based cognitive radio network. *IEEE Commun Lett* 22(8):1708–1711
10. Zheng Y, Zhiguo D, Pingzhi F, Naofal AD (2016) A general power allocation scheme to guarantee quality of Service in Downlink and Uplink NOMA systems. *IEEE Trans Wirel Commun* 15(11):7244–7257
11. Nibedita N, Sudhan M, Hsiao-Chun W (2018) Maximizing secrecy capacity of underlay MIMO-CRN through bi-directional zero-forcing beamforming. *IEEE Trans Wirel Commun* 17(8):5327–5337

Chapter 31

Hybrid Optimization of Cuckoo Search and Differential Evolution Algorithm for Privacy-Preserving Data Mining



J. Sudeeptha and C. Nalini

Abstract Data mining applies data analysis techniques to find patterns as well as relations in information to make good decisions. In today's digital world, preserving data privacy is a challenging task due to business enterprise applications' leverage to modern technologies. The k-anonymity model is used to protect each single record from being identified by making all records indistinguishable from k-1 other records. Suppression and generalization methods are used to implement k-anonymity. In this chapter, generalization method is used to implement k-anonymity. To improve the accuracy of a classification algorithm, hybrid optimization of Cuckoo Search (CS) with Differential Evolution (DE) search technique is used to select optimized generalized feature set. The classification accuracy is evaluated using public dataset with and without k-anonymity.

Keywords Data Privacy · k-anonymity · Generalization · Cuckoo Search (CS) · Differential Evolution (DE) · Classification

Abbreviations

CS	Cuckoo Search
PPDM	Privacy-preserving data mining
DE	Differential Evolution
QID	quasi-identifier

J. Sudeeptha
Department of Computer Science Engineering, Kongu Engineering College, Erode, Tamil Nadu, India

C. Nalini (✉)
Department of Information Technology, Kongu Engineering College, Erode, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_31

31.1 Introduction

Privacy-preserving data mining (PPDM) uses data mining techniques to protect data privacy during mining process. Most PPDM methods transform original data to perform privacy preservation. Randomization utilizes data distortion techniques for adding noise to original information. Retrieving the original values is difficult, but merely aggregated distribution is retrieved. The two types of perturbation with randomization are additive perturbation and multiplicative perturbation. In additive perturbation, randomized noise is appended to information records. In multiplicative perturbation, arbitrary projections or arbitrary rotation modes are used to perturb records. Generalization and suppressions anonymization techniques are used for making individual records indistinguishable among a set of records. The k -anonymity model protects individual records from identification and works through making all data records indistinguishable from $k-1$ other records; k -anonymity model protects released information from potential reidentifications of individuals to whom the data refers. In an intuitive manner, k -anonymity says that all data releases must ensure that all values combinations of released features are externally accessible as well as capable of being exploitable to a link that indistinctly matches to a minimum of k -individuals. K -anonymous data mining is a recent introduction to ensure preservation of privacy during the release of data mining outcomes.

The protection ensures the likelihood of figuring out individuals on the basis of released information in dataset that is not more than $1/k$. Generalizations substitute feature values with semantically accurate though less detailed values. Suppression removes certain attribute values substituting appearances of values with a special character “?” revealing that any value may be utilized in its place. k -anonymity method masks sensitive data and makes it hard to re-identify “Let $RT (A_1 \dots A_n)$ be a table and QRT be a related quasi-identifier.” RT satisfies k -anonymity only if each value sequence in RT [QRT] appears minimum k times in RT [QRT]. Classical optimization methods are useful to locate continuous as well as differentiable function’s optimum solution. The analytical approaches utilize techniques of differential calculus to locate optimal points.

A dataset contains relevant and irrelevant attributes to a task. Selecting an optimized relevant attribute from a dataset is a challenging task. Various optimization methods are used in this area. Particularly, hybrid optimization search techniques are used for selecting optimal generalized feature set. This work suggested hybrid optimization is based on Cuckoo Search with DE. An adult dataset is used to evaluate classification accuracy with anonymization.

31.2 Related Work

A novel method wherein requisites on distortion amounts permitted on original data were enforced to retain utility was considered by Campan et al. [1]. The approach consisted of delineating quasi-identifiers' generalization constraints attaining p -sensitive k -anonymity within enforced constraints. Experiments showed the proposed algorithm to be comparable to existing algorithms to generate p -sensitive k -anonymity regarding results' quality. Obviously obtained masked microdata comply with generalization constraints as indicated by users.

Fung et al. [3] surveyed, summarized, and evaluated different PPDP approaches studying problems in practical data publishing and clarifying differences as well as requisites distinguishing PPDP from other issues. They also proposed further directions research ought to take. An optimum balance between privacy and utility while publishing any organization's dataset was found by Jena et al. [4]. K-mean algorithm was used to cluster the dataset followed by k -anonymity.

A practical approximation model resolving the k -anonymity issue with approximation guarantee was presented by Tassa [5] that achieved reduced information losses than leading approximation algorithm and leading heuristic algorithms. The proposed algorithm's modified version that provided ℓ -diverse k -anonymity achieved lesser data loss than related leading algorithms altered variations.

A novel technique to achieve k -anonymity called k -anonymity of Classification Trees Using Suppression (kACTUS) was proposed by Kisilevich et al. [6]. Effective multidimensional suppressions were carried out in kACTUS, that is, values were suppressed in specific records on the basis of other feature values, with no manually yielded domain hierarchy trees. The results proved that predictive performance of kACTUS was better than current k -anonymity models. Lindell and Pinkas [7] surveyed secure multiparty computation basic paradigms and notions and discussed efficiency issues demonstrating difficulties in construction of extremely effective models.

A hybrid EA utilizing GA as well as PSO was implemented by Mandapati and Chekka [8]. Both GA and PSO work with the same population. k -anonymity was attained by original database generalization in the suggested model. The hybrid optimization searched for the best generalized attribute set. Anonymization of query logs through usage of micro aggregations was presented by Navarro-Arribas et al. [9] which ensured users k -anonymity in the query log as well as maintaining usefulness of data. It evaluated the proposed investigation in real query logs, showed privacy as well as utility attained, and provided predictions for usage of data in clustering-based data mining procedures.

Various anonymization techniques used for privatizing data were discussed by Sumana and Hareesha [10]. Anonymization's goal is in securing access to confidential information while simultaneously displaying aggregated data to the public. Each technique's challenge is data protection to enable publishing with no revelation of private data related to particular persons. Furthermore, security must be attained with least accuracy loss looked for by database clients. Different anonymization approaches were discussed, and a comparison of the same was provided.

31.3 Methodology

31.3.1 *The Adult Dataset*

UCI Machine Learning Repository provided "Adult" dataset for evaluator purposes. It had 48,842 samples, with both categorical and integer features. Privacy-preserving classification models are tested solely on Adult datasets that are now a common benchmark for k-anonymity. Adult datasets possess six numeric as well as eight categorical features. The target feature is income level with two potential values, "<=50 K" or ">50 K." The dataset consists of the following fields:

- Age: continuous, 2. Workclass: 8 values, 3. Fnlwgt: continuous, 4. Education: 16 values, 5. Education-num: continuous, 6. Marital-status: 7 values, 7. Occupation: 14 values, 8. Relationship: 6 values, 9. Race: 5 values, 10. Sex: Male, Female, 11. Capital-gain: continuous, 12. Capital-loss: continuous, 13. Hours-per-week: continuous, 14. Native-country: 41 values, 15. >50K Income: Yes, No.

31.3.2 *k-anonymity*

The k-anonymity algorithm differentiates three objects including individuals whose privacy is to be protected, dataset owners who control tables where all rows describe a single person, and attackers. The k-anonymity algorithm assumes two things:

Under k-anonymity algorithm, dataset owners retain individuals' k-anonymity if none of them are connected with less than k rows in a divulged table. This is accomplished through ensuring that in any owner-divulged table, there are a minimum of k rows with the same combination of values in public features. As that is not true for all tables, most research under k-anonymity algorithm [2] focuses on methods to suppress, alter, and eliminate attribute values so that the altered table qualifies as k-anonymous.

The k-anonymity algorithm presumes a quasi-identifier (QID), a group of features which function as identifiers in dataset. In a simple instance, the assumption is made that datasets are tables with all tuples corresponding to individuals. For attaining k-anonymity, certain attributes values in quasi-identifiers are generalized through replacement of values in lower levels by values at higher levels based on generalization hierarchies.

J48 algorithm based on decision tree is a basic C4.5 decision tree for classification creating binary trees. When tree building, J48 does not consider missing values, that is, value for an entity that can be estimated on the basis of that feature values knowledge about other tuples.

Algorithm [10] J48:

```

INPUTS:
    D //Training information OUTPUTS
    T //Decision trees
DTBUILD (*D)
{
    T=φ;
    T= Generate root node and label with splitting feature;
    T= Append arc to root nodes for every split
    predicate and label;
    For every arc do
        D=Databases generated by application of splitting predicate to D;
        If termination criterion is arrived at for the route, then
            T'= generate leaf node and label with suitable class;
        Else
            T'= DTBUILD(D);
            T= append T' to arc;
    }

```

31.3.3 Cuckoo Search (CS)

A meta-heuristic approach, Cuckoo Search is inspired by obligating brood parasitic activity of a bird species known as Cuckoos [11]. The birds leave their eggs in host birds' nests with great capacity to select recently created nests while also discarding already present eggs, thereby increasing their eggs' hatching probability. The host bird looks after these eggs assuming them to be its own.

CS has its basis in three rules:

- All cuckoos lay a single egg at a time, randomly dumping it in a selected nest.
- The best nests with the greatest quality of eggs (solutions) are carried over to the subsequent generation.
- Quantity of accessible host nests is set, and hosts discover alien eggs with probability $p_a \in [0, 1]$. Here, host birds either throw away the eggs or abandon the nest, building an entirely fresh nest in another place.

31.3.4 Differential Evolution

DE is a comparatively recent heuristic formulated for optimizing issues on continuous fields where decision parameter is denoted by real numbers in the vector. As in other evolutionary algorithms, DE's initial population is randomly generated and evaluated. Then mutation takes place, during which three parents are selected, generating one offspring competing with a parent to decide who passes on to the next generation. New individuals are generated by DE through linear combination of two or more individuals already in the recent population by executing a particular operator [12]. Present operators define distinct DE versions; operators are sorted as per notation DE/x/y/z

Wherein,

- x – delineates vector to undergo mutation which can be either RAND, implying arbitrary selection from the entire population, or BEST, implying that the best known solution is utilized.
- y – refers to the quantity of difference utilized by vectors (1 or 2).
- z – represents type of likelihood of carrying out linear combination are drawn; one may presume that either value BIN corresponds to independent binomial experiments for all dimensions or value EXP corresponds to conditional probability for every dimension with respect to the earlier one.

The proposed hybrid optimization looks for best generalized attribute set. The proposed hybrid optimization is based on Cuckoo Search (CS) with Differential Evolution (DE). The DE mutation operator adds diversity of the population to improve search efficiency. Following are the steps in the suggested hybrid CS-DE model:

```

Input : the feature set
/Initialize Cuckoo search/
Generate a population of n host nests
   $x_i, i=1,2,3,\dots,n$ 
for all  $x_i$  do
  Compute fitness  $F_i = f(x_i)$ 
end for
while (Number Objective Evaluations < Max Number Evaluations) do
  Create a cuckoo egg ( $x_j$ ) by taking Lévy flights from arbitrary nests
   $F_j = f(x_j)$ 
  Select an arbitrary nest  $i$ 
  If ( $F_j < F_i$ ) then
     $x_j = x_i$ 
     $F_i = F_j$ 
  end if
  Discard a fraction  $p_a$  of worst nests.
  Construct fresh nests at fresh locations via Lévy flights for replacing nests
  which are lost.
  Assess fitness of fresh nests and rank all solutions.
  /Initialize Differential Evolution/
  The nests from the cuckoo search are used as the initial population
  for  $i= 1$  to  $N$  do
    compute a mutant vector  $u$ ;
    crossover to  $u$  and  $x_i$  to form  $y$ 
    if  $f(y) < f(x_i)$  then replace  $x_i$  with  $y$ 
  end for;
  until stopping condition
end while

```

31.4 Results and Discussion

The experiments are conducted without optimization, with Cuckoo Search optimization and with hybrid search optimization for no anonymization and for the values of $k = 5, 20, 35, 50, 65, 80$. J48 classification algorithm is used to evaluate the accuracy. Table 31.1 and Fig. 31.1 illustrate the classification accuracy for various k values.

Table 31.1 Classification Accuracy

k-anonymity size	Without optimization	With Cuckoo Search optimization	With hybrid Cuckoo Search
No anonymization	85.73	85.73	85.73
K=5	85.33	85.41	85.89
K=20	82.88	84.59	84.84
K=35	81.65	84.02	84.79
K=50	81.49	81.97	82.38
K=65	81.31	81.77	82.16
K=80	80.54	81.61	81.77

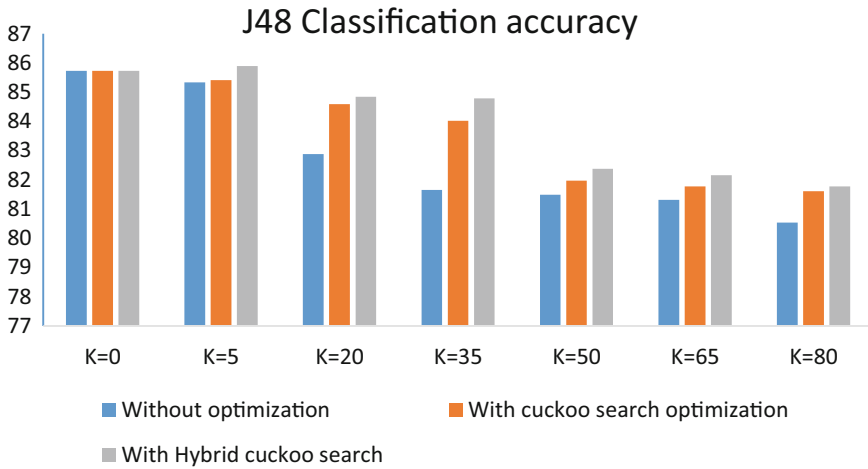


Fig. 31.1 Classification accuracy

Table 31.2 RMSE

k-Anonymity size	Without optimization	Cuckoo Search	Hybrid Cuckoo Search with DE optimization
No anonymization	0.3253	0.3253	0.3253
K=5	0.3279	0.32012877	0.317311637
K=20	0.3956	0.39018028	0.387917234
K=35	0.3988	0.38536044	0.381198547
K=50	0.4102	0.40868226	0.405494538
K=65	0.4124	0.41128652	0.408901058
K=80	0.4176	0.40770288	0.405338203

The hybrid approach shows a better performance compared with the other techniques when k=35, and all the techniques produced the same result compared with other techniques when there is no anonymization. Table 31.2 shows the results of RMSE for the different k values.

Table 31.2 shows the results of RMSE for the different k values for all the techniques used in this experiment. The algorithm converged when $k=35$. The Cuckoo Search with differential evolution algorithm helps to select the optimized generalized feature set and also enhance the accuracy of the classification algorithm.

31.5 Conclusion

The results reveal that the proposed hybrid approach, i.e., Cuckoo Search (CS) with Differential Evolution (DE) k -anonymity model, is a promising method to preserve and protect sensitive values of record's features from attackers and also to increase the accuracy of a classification algorithm.

References

1. Campan A, Truta TM, Cooper N (2010) P-Sensitive K-Anonymity with Generalization Constraints. *Trans Data Priv* 3(2):65–89
2. Friedman A, Wolff R, Schuster A (2008) Providing k -anonymity in data mining. *VLDB J* 17(4):789–780
3. Fung B, Wang K, Chen R, Yu PS (2010) Privacy-preserving data publishing: a survey of recent developments. *ACM Comp Surv* 42(4):14.2–14.52
4. Jena L, Kamila NK, Mishra S (2013) Optimizing the convergence of data utility and privacy in data mining. *Int J* 2(1):155–166
5. Kenig B, Tassa T (2012) A practical approximation algorithm for optimal k -anonymity. *Data Min Know Disc* 25(1):134–168
6. Kisilevich S, Rokach L, Elovici Y, Shapira B (2010) Efficient multidimensional suppression for k -anonymity. *IEEE Trans Know Data Eng* 22(3):334–347
7. Lindell Y, Pinkas B (2009) Secure multiparty computation for privacy-preserving data mining. *J Priv Confiden* 1(1):5–39
8. Mandapati S, Chekka RB (2013) A hybrid algorithm for privacy preserving in data mining. *Int J Intel Syst Appl* 5(8):47–53
9. Navarro-Arribas G, Torra V, Erola A, Castellà-Roca J (2012) User-anonymity for privacy preserving data mining of query logs. *Info Process Manage* 48(3):476–487
10. Sumana M, Hareesha KS (2010) Anonymity: an assessment and perspective in privacy preserving data mining. *Int J Comp Appl* 6(10):1113–1457
11. Valian E, Mohanna E, Tavakoli S (2011) Improved cuckoo search algorithm for global optimization. *Int J Comm Info Tech* 2(3):36–44
12. Wang G, Guo L, Duan H, Liu L, Wang H, Wang B (2012) A hybrid heuristic DE/CS algorithm for UCAV path planning. *J Info Comp Sci* 5(16):4811–4818

Chapter 32

Using Sliding Window Algorithm for Rainfall Forecasting



M. Vijaya Chitra and Grasha Jacob

Abstract Rainfall forecasting has been an onerous task to deal with. However, it is a part and parcel to sustain our life since it affects not only the agriculture growth but also the farming community. Rainfall prediction will definitely pose a great challenge but for meticulous planning and management of water resources. Therefore, this chapter presents an approach for rainfall prediction using Sliding Window concept with Jaccard distance metric measure. The Sliding Window Algorithm watches the information during a similar period in an earlier year and predicts precipitation in the next year. Using Sliding Window Algorithm, the precipitation expectation test was tested for Tirunelveli District, Tamil Nadu, India, using the rainfall data for a 10-year period.

Keywords Rainfall forecasting · Sliding window · Root-mean-square error · Jaccard distance

Abbreviations

SWA	Sliding window algorithm
MSE	Mean square error
GRMSE	Geometric root-mean-square error
RMSE	Root-mean-square error
JD	Joint director
EY	Earlier year
PY	Present year
EV	Earlier variation
PV	Present variation
MPV	Mean present variation

M. Vijaya Chitra (✉) · G. Jacob

Department of Computer Science, Rani Anna Government College for Women, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_32

MEV	Mean earlier variation
PRV	Predicted variation
AR	Average rainfall
FR	Forecasted rainfall

32.1 Introduction

Scientific method of rainfall prediction has come to stay, replacing the traditional method. Farmers in India most probably follow the traditional method, unlike the method prevailing in Israel. For instance, drip irrigation is scrupulously followed in Israel, whereas in India, it is followed then and there. Basically our country is subtropical. Water dependence is monsoon related, and it has a toll on our production in farming. Besides, more number of people are employed in agricultural field. But the production is inversely proportional. It is evident that the production is not commensurate to the people employed.

There are three important things for sustenance of life: food, clothing, and shelter. As for food, India is primarily an agricultural-based country. So rainfall and underground water are inter-related. Climate forecasting requires a scientific bent of mind. However, people have made their attempts to predict the climate casually. This paves way for the most established human progress.

There are two models that rule over as for the calculation of rainfall: empirical and dynamic methods. Physical models based on the system of equations play a vital role in the dynamical approach, whereas the empirical approach is based on analysis of past historical data of weather [1]. In this study, the sliding window idea is proposed so as to create acceptable precipitation estimating result.

Datar et al. [2] suggested a model of estimating specifically sliding window. In the information mode measurements, as for the sliding window, the setting of N has taken over. The result obtained for a solitary stream was extracted from evaluating capacities. A fundamental issue that they considered is to decide the quantity of 1's in a sliding window, which is known as the Basic Counting issue.

Kapoor and Singh [3] had proposed sliding window calculation (SWA) to forecast climate condition. In order to check the exactness, they experimented month-wise results for a long time. In their examination, they arrived at a conclusion that their method had been found to be most accurate except for the months of seasonal change where conditions are highly unpredictable.

Rajini Kanth et al. [4] implemented the k-clustering method to gather similar datasets to predict rainfall. Using linear regression, it would need higher logical techniques like machine learning algorithms. Using this, effective investigations and predictions of climate conditions have been done.

Viswambari and Anbu Selvi [5] suggested the data mining techniques to forecast rainfall, wind speed, and moisture. They survey the various techniques implemented in data mining to predict weather.

Azahari, Mahmood, and Rizauddin [6] had executed the sliding window idea that was proposed by Kapoor and Sarabjeet, keeping in mind the end goal to deliver acceptable exactness of precipitation anticipating result. The result shows that the upgraded sliding window estimation is deeply exact. However it contrasted with the past sliding window estimation, as a result of model approval utilizing mean square error (MSE) and relative geometric root-mean-square error (GRMSE).

In this examination, the sliding window idea has been created with a specific end goal to deliver estimated result by characterizing the normal month-to-month precipitation, from January 2008 to December 2016. In this work, Jaccard distance metrics is used in SWA. This distance metrics is calculated to find the closeness for predicting rainfall.

In this study, the dataset used belongs to Joint Director (JD) Office, Tirunelveli District, Tamil Nadu, India. This is a district where farming is essential. Both food crops and commercial crops are grown in the district.

Tirunelveli district is privileged with the Western Ghats from which all the perpetual streams and rivers flow towards the east. Precipitation is essential for the constant development of this region, and hence the rainfall in this district is analyzed for forecasting. The data used here for testing was collected from the JD Office, Tirunelveli, Tamil Nadu. The data covers the period of 108 months, that is, from January 2008 to December 2016.

32.2 Proposed Work

In this study, the sliding window concept augmented the forecasting result by defining the average monthly rainfall, from January 2008 to December 2016. Using Sliding Window Algorithm with Jaccard distance, month-wise results are being computed to check the exactness.

32.2.1 Methodology

Climate conditions are subject to change. In this work, a philosophy is being suggested that could scientifically demonstrate these two varieties and utilize them to foresee the future precipitation. For this, the climatic conditions for the previous 2 years and present 1 year will be considered. The block diagram of the methodology is shown in Fig. 32.1.

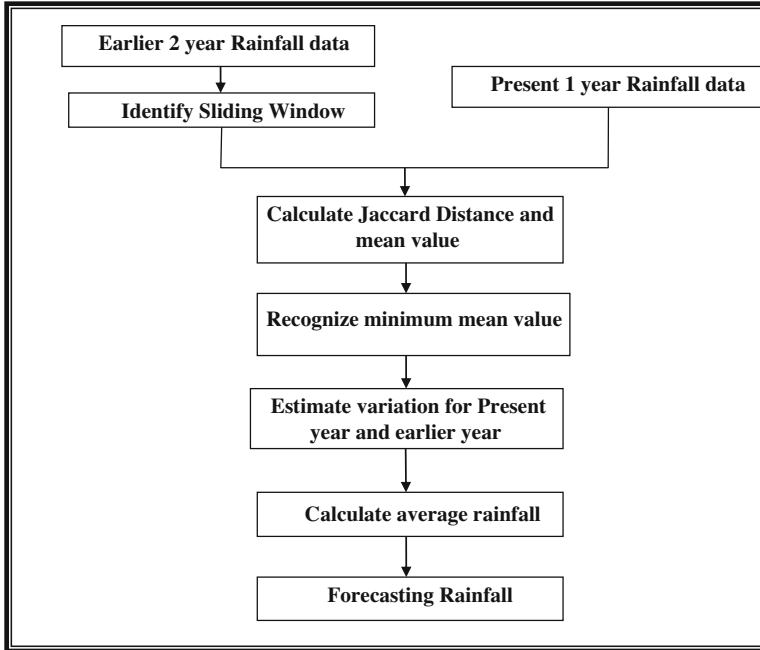


Fig. 32.1 Block diagram

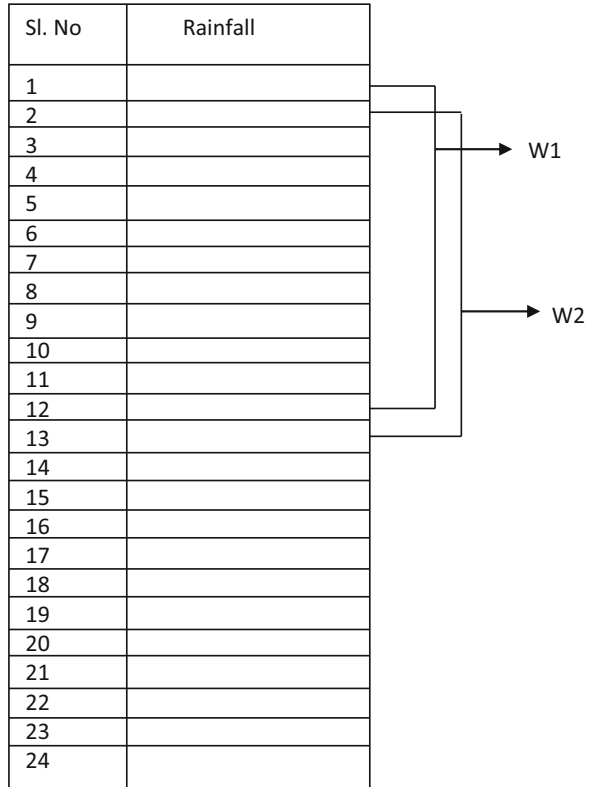
For instance, if the precipitation of the year 2013 is to be foreseen, then consider the conditions from two earlier years 2010 and 2011. In order to model the above dependencies the current year's variation is being matched with those of previous years by making use of sliding window. The best planned result from the sliding window is taken. The picked window and the present year 2012 are utilized to predict the precipitation of the year 2013.

32.2.2 *Sliding Window*

In this work, only one weather parameter will be taken into consideration, that is, monthly rainfall. Hence the current year will be represented by a matrix of size 12×1 for 1 year. Similarly for the previous 2 years, the matrix size would be 24×1 . Next divide the matrix of size 24×1 into the sliding windows.

The concept of sliding window is shown in Fig. 32.2. W1 represents window1 and W2 represents window2. Hence, 13 sliding windows can be made of size 24×1 each.

Fig. 32.2 Sliding window for 24 months



32.2.3 Statistical Measures

32.2.3.1 Mean

Mean is the average of the total climate information of all the years considered.

32.2.3.2 Variation

Variation gives the distinction between the earlier year's climate and current year's climate.

32.2.3.3 Jaccard distance

It calculates the similarity between two dissimilar sets:

$$|x \cap y| / |x \cup y|$$

32.2.3.4 Root-Mean-Square Error

Root-mean-square error (RMSE) is a frequently used measure of the contrast between the predicted value and the original value:

$$\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}$$

32.2.4 Sliding Window Algorithm

The initial step of this calculation is to build the present year matrix and earlier year matrix. The matrix size is 12×1 and 24×1 . Twelve months for present year (PY) and 24 months for earlier year (EY) were taken. Then it is identified by 13 sliding windows from the earlier year matrix (EY). Now the size of the previous year matrix is 12×1 . The sliding window algorithm using Jaccard distance for predicting the monthly rainfall is shown in Algorithm 1.

In the next step, computation of Jaccard distance has been done. Then the mean value is found, and the minimum mean value for the sliding window is selected. The variation for earlier year (EY) and present year (PY) is estimated. Then the mean values for the above variations are calculated. Based on the mean values, the predicted variation (PRV) is proceeded. Next, average rainfall (AR) is measured. Finally the forecasted rainfall (FR) is evaluated.

Algorithm 1: Sliding Window Algorithm Using Jaccard Distance

Step 1: Take 12×1 matrix as present year (PY) and 24×1 matrix as earlier year (EY).

Step 2: Identify 13 sliding windows from earlier year matrix (EY).

Step 3: Compute Jaccard distance for sliding window and find the mean value.

Step 4: Choose the minimum mean value, from window.

Step 5: Compute variation PV for minimum mean value for present year (PY).

Step 6: Compute variation EV for earlier year (EY).

Step 7: Calculate mean value for the variation.

$$\text{MPV} = \text{Mean (PV)}$$

(continued)

Algorithm 1 (continued)

$$\mathbf{MEV} = \text{Mean (EV)}$$

Step 8: Compute predicted variation, $\mathbf{PRV} = \frac{\mathbf{MPV} + \mathbf{MEV}}{2}$

Step 9: Calculate average monthly rainfall, **AR**.

$$\mathbf{AR} = \frac{\sum \text{rainfall for } n \text{ years}}{n}$$

Step 10: Forecast rainfall

$$\mathbf{FR} = \mathbf{PRV} + \mathbf{AR}$$

32.2.5 Dataset Description

In this work, the Tirunelveli District rainfall data from the year 2008 to 2017 were considered. For using sliding window algorithm, the data for the years 2014–2016 were examined. From this, the data for the year 2017 were calculated and verified. Finally, from the above outcome, the rainfall data for the year 2018 is forecasted.

32.3 Results and Discussion

Table 32.1 depicts the analysis results obtained in millimeters and root-mean-square error for the year 2017 using Jaccard distance metric measures. Fig. 32.3 exhibits the 2017 forecasting results. The graphs indicate that the test results are in accordance with the nature of the rainfall received in 2017. The precipitation result of the months August, September, October, and November was better compared to the other months.

Table 32.2 shows the prediction results in millimeters for the year 2018. Fig. 32.4 shows the 2018 forecasting results. Using Jaccard distance, the rainfall information for the year 2018 is calculated. Based on the outcome, the Jaccard distance produced **30.04** RMSE value.

Table 32.1 Prediction results for the year 2017

Month	Original	Jaccard Distance
January	62.51	42.86
February	4.66	44.66
March	60.66	98.87
April	27.09	74.10
May	80.85	67.07
June	46.68	60.21
July	24.92	46.22
August	40.39	46.74
September	64.48	56.57
October	282.98	260.56
November	334.41	302.42
December	231.36	179.18

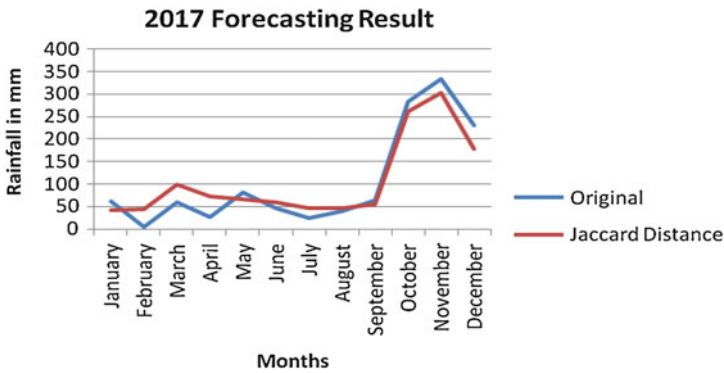


Fig 32.3 Prediction results for the year 2017

32.4 Conclusion

In this chapter, the simplest yet viable technique Sliding Window Algorithm has been applied for the rainfall data to forecast rainfall in Tirunelveli District, Tamil Nadu, India. This research will offer a helping hand in assisting the farmers to know how much quantity of rainfall will occur in the forthcoming season, so that they could mitigate the problems arising in the crop pattern. As a consequence, the farmers will definitely derive maximum benefits.

Table 32.2 Forecasting results for the year 2018

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
	59.26	46.41	74.18	76.11	86.00	74.74	55.57	53.72	71.93	264.09	329.72	204.84

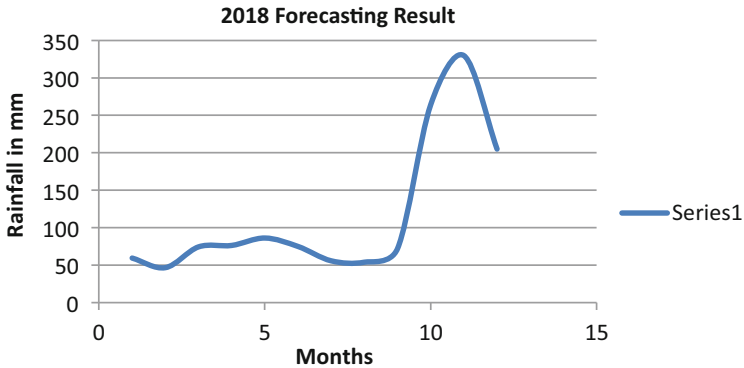


Fig 32.4 Forecasting results for the year 2018

References

1. Hirani D, Mishra N (2016) A survey on rainfall prediction techniques. *Int J Comp Appl* 6 (2):28–42
2. Datar M, Gionis A, Indyk P, Motwani R (2002) Maintaining stream statistics over sliding windows. *SIAM J Comp* 31(6):1794–1813
3. Kapoor P, Bedi SS (2013) Weather forecasting using sliding window algorithm. Hindawi Publishing Corporation, *ISRN Signal Processing*, 2013, Article ID 156540.
4. Rajini Kanth TV, Balaram VVSSS, Rajasekhar N (2014) Analysis of Indian weather data sets using data mining techniques, pp. 89–94, © Cs & It-Cscp 2014
5. Viswambari M, Anbu Selvi R (2014, June) Data mining techniques to predict weather: a survey, *ISSN 2348–7968, IJISSET-Int J Innov Sci Eng Technol*, 1(4)
6. Azahari F, Othman M, Saian R (2017) An enhancement of sliding window algorithm for rainfall forecasting. In 6th international conference on computing and informatics, 25–27

Chapter 33

Air Pollution-Level Estimation in Smart Cities Using Machine Learning Algorithms



M. Nelgadevi and Grasha Jacob

Abstract Air pollution is a serious issue that has been harming and hurting the people in India. The air is contaminated owing to industrial plants and manufacturing activities, combustion from fossil fuels, farming chemicals and household products and natural events like volcanic eruptions, forest fires and gaseous releases from decaying plants and animals. Air pollution not only harms the comfort and health of both humans and animals but also destroys the life of the plants. Air pollution is otherwise called as environmental pollution that causes serious problems confronting humanity and other life forms on planet Earth today. In this work, K-nearest neighbour method is used to evaluate the position of air pollution at several places in Chennai City. Random Forest and Support Vector Machine algorithms evaluate the efficiency of the proposed model, thereby categorising the data into six classes of pollution levels.

Keywords Environmental pollution · Air quality · K-nearest neighbour · Random forest · Support vector machine

Abbreviations

COPD	Chronic obstructive pulmonary diseases
AQI	Air quality index
EPA	Environmental protection agency
RBF	Radial basis function
NARX	Nonlinear auto regressive models with exogenous inputs
SVM	Support vector machine
K-NN	K-nearest neighbour
RF	Random forest

M. Nelgadevi (✉) · G. Jacob
Department of Computer Science, Rani Anna Government College for Women, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu, India

RSPM	Respirable suspended particulate matter
CPCB	Central pollution control board

33.1 Introduction

The contamination of earth's environment known as pollution is endangering human health, quality of life and the natural functioning of the ecosystems. This is further aggravated due to urbanization, development of many industries, absence of awareness, poor maintenance of automobiles, damaged road conditions, etc. It is one of the largest potential killers in India. Day to day, the death toll increases. The people in India are prone to chronic obstructive pulmonary diseases (COPDs), heart diseases, lung cancer and so on. The premature deaths are projected to go up owing to outdoor air pollution affecting three million people globally in 2010 to a global total of six to nine million people in 2060. This is further jeopardising due to heavy vehicular traffic and industrial emissions. So, air automatically gets polluted with the mixture of gases and has an adverse effect on health which is a major issue worldwide. The air pollutants with corresponding source and effects are listed in Table 33.1.

The most effective way to deal with air pollution is societal participation. The quality of air has been posing a challenge to the human community all over the world. So, one should be alert of the quality of air inhaled. It is relevant that we should arrive at accurate methods to measure the air quality in advance. The public has every right to know the details of air quality, even a layman should understand. Air quality index (AQI) is one such active tool for the supply of air quality information to the public.

Air Quality Index

An air quality index (AQI) is a number used by [government agencies](#) to make it known to the public regarding [air pollution](#). It is a tool to highlight the status of air quality. The complex air quality data of different pollutants appear to be a single number and colour. AQI has six classes of air quality level. These are Good, Satisfactory, Moderately Polluted, Poor, Very Poor and Severe. Each of these categories mentioned causes impacts on human health and the environment. The increase in AQI is directly proportional to the adverse health effects on human beings. The air monitor examines the air pollutant concentration over a specified averaging period for estimation of air quality. Environmental Protection Agency (EPA) evaluates the AQI for obtained air contaminants such as particulate matter, carbon monoxide, ammonia, sulphur dioxide, nitrogen dioxide and ozone. The AQI was categorised based on the assessment of these pollutants (PM₁₀, PM_{2.5}, CO, NH₃, SO₂, NO₂ and ozone). Pollutant concentrations are transformed into AQI which are assumed to be in the range 0–500. The aim of the AQI is to enable laypersons to realise quality of air and its associated impacts on human health. Air quality index values are typically grouped into ranges and colour code and identical public health advisory is exhibited in Table 33.2.

Table 33.1 Source and effects of air pollutants

Air pollutants	Source and effects
Carbon monoxide	Formed from incomplete combustion of carbon-containing fuels.
	Largest human activity that leads to the formation of CO is vehicle emissions.
	Inhaling high concentrations of CO leads to reduced oxygen and causes health illness like headaches and increased risk of chest pain with heart disease
Nitrogen dioxide	Produced as an outcome of road traffic and other fossil fuel combustion practices
	Inhaling raised levels of nitrogen dioxide is the major reason for respiratory issues
Particulate matter	Combination of solids and liquid droplets moving in the air
	Particles less than or equal to 10 and 2.5 $\mu\text{g}/\text{m}^3$ in diameter can penetrate into the lungs, causing serious health issues like high blood pressure, heart attack and strokes and in serious conditions, leads to premature death
Ammonia	Major source of emissions is agriculture, which includes animal farming and NH_3 -based application which uses fertilizer.
	Inhaling high concentrations of ammonia in air causes immediate irritation in the eyes, nose, throat and respiratory tract and also leads to blindness and lung damage
Sulphur dioxide	Gas which is the outcome of fossil fuel combustion at power plants, other industrial facilities, as well as in transports.
	It creates severe infection in the nose, throat and chest and leads to coughing, wheezing, breathing issues, etc.
Ozone	Produced through composite series of photochemical reactions of sunlight and heat.
	Exposing to ozone causes severe injury to lungs. Inhalation of less quantity causes chest infection, cold and cough with breathing problems
Carbon dioxide	Formation of the gas from both human and natural process
	Human process from actions like cement manufacture, deforestation, boiling of fossil fuels like coal, oil and natural gas
	Destroys human health by relocating oxygen in the atmosphere which leads to severe suffocation with headaches, unconsciousness, etc.

Table 33.2 AQI with six classes of corresponding colour codes

Air Quality Index (AQI)	Associated Health Effects	Indicated by hecolours
Good (0–50)	Impact is less and little risk	Green
Satisfactory (51–100)	Minor respiratory discomfort acts on sensitive people	Light Green
Moderately Polluted (101–200)	Harmful to people with asthma, heart disease, risk for children and elderly people.	Yellow
Poor (201–300)	Breathing discomfort to heart patient on prolonged exposure	Orange
Very Poor (301–400)	Causes respiratory illness to people with lung and heart diseases	Red
Severe (401–500)	Causes respiratory problems even on healthy people; impacts may be experienced by the entire population	Maroon

AQI value is computed from the pollutant concentrations, which results in a value that may be commonly taken by the public for understanding about air quality to know whether status of the air is good or bad. When AQI values are beyond 100, air quality is unhealthy for certain sensitive groups of people first and then for everyone as AQI increases.

33.2 Literature Survey

For the past several years, the quality of air is estimated using several approaches with health concern.

Dragomir [3] used the class of the k-nearest neighbour to categorise the pollution level to forecast the air quality in order to predict the value of the air quality index.

Prakash Mamta and Bassin [6] observed that the calculated AQI values for SO₂ and NO₂ are categorised under ‘good’ and ‘good-to-moderate’. It was found as ‘poor’ and ‘extremely poor’ for RSPM and SPM.

Chi-Man Vong et al. [1] constructed the suitable choice of kernel in Support Vector Machine for predicting ambient air pollutant. The choice of Linear and Radial Basis Function (RBF) model was relatively good in the prediction of SO₂ and NO₂. In seasonal test, these models produced superior results with relatively lower errors when compared with other models of SVM.

Yin Zhao and Yahya Abu Hasan [7] built PM_{2.5} concentration level predictive models based on ANN, Boosting (i.e. AdaBoost.M1) and k-NN. Ada Boost model has accuracy higher than K-NN model.

L. Wang and Y. P. Bai [5] adopted the Nonlinear Auto Regressive models with Exogenous Inputs (NARX) and Support Vector Machine (SVM) for prediction of the air quality index. NARX is more appropriate than the SVM in predicting the AQI.

Chuanting Zhang and Dongfeng Yuan [2] proposed a fast fine-grained AQI level prediction method based on the implementation of random forest algorithm on Spark. The method is fast in predicting concentration level of PM_{2.5}.

H. Wang et al. [4] applied fuzzy time series theory to forecast air quality index and found that its prediction is better than the traditional time series forecasting model.

33.3 Proposed Model

It is required that the classification of air quality has to be estimated in real time and the projected model is based on three different classifiers of machine learning. One-year data from Chennai is chosen to decide on the level of pollution for the particular area. The K-Nearest Neighbour, Random Forest and Support Vector

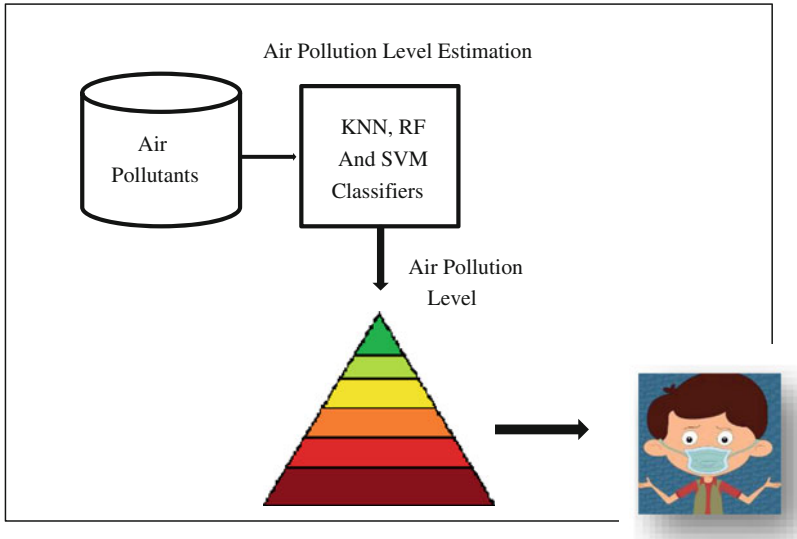


Fig. 33.1 The architecture of air pollution estimation model

Machine classifiers are related to estimate the classification accuracy. The design of the system is diagrammatically symbolised in Fig. 33.1.

33.3.1 Classification of Air Quality Level

33.3.1.1 K-Nearest Neighbours

In this study K-Nearest Neighbours (K-NN), a Machine Learning Algorithm, is used for the classification of multi-class of pollution level. K-NN is based on the knowledge that it classifies a new data based on attributes and samples of training dataset, using a majority of K-nearest neighbour group. Hence, it does not essentially attain a model from training dataset but store the data. Given a new data, its class label is obtained by k-nearest neighbours of the stored training data. $K = 3$ contributes highest accuracy for this model.

The Euclidean distance is used as the distance metric to compute the k-nearest neighbours and is calculated as

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (33.1)$$

The nearest neighbours are the training data that are close to the test data that helps to predict the class. Based on the majority vote of the K-nearest neighbours, the test data is classified to a particular class of pollution level.

Algorithm K-Nearest Neighbour

1. Load the training dataset
 2. Choose the K value fit to this model
 3. For getting class label for test data
 - compute the Euclidean distance to all data in training set
 - sort the Euclidean distances in ascending order
 - get the first k values as nearest neighbours
 - assign a class to the test data based on the majority of classes given by the k-nearest neighbours
- End

33.3.1.2 Random Decision Forest

The Random Decision Forest is a model which operates by constructing several decision trees in the training stage. The decision of the majority of the trees is taken by the random forest algorithm as a final output. In this study, the training dataset of the pollutants concentration are distributed into bootstrap subsets to construct number of decision trees. The gini index measure to select best splitting attributes is used to build the decision trees. The decision tree is constructed until the leaf node contains the class label indicating the pollution level. The decision attained by all the trees is assembled to yield the final output. The training dataset is used to build the decision tree, and the result obtained by the decision tree is used to find the status of air quality. The testing dataset is utilised to test the model.

33.3.1.3 Support Vector Machine

In this study, Support Vector Machine (SVM) algorithm is used for classification of pollution data to estimate the pollution level. The training dataset are mapped to high dimensional space to acquire the nonlinear class boundaries to implement a multi-class classification. The kernel acts as a hyper-plane which linearly separates the dataset with corresponding class labels indicating the status of air pollution. The technique to separate the classes as wide as possible is done by Radial Basis Function. The data points nearest to the hyper-plane are the support vectors to classify the data in an efficient way. The regularisation parameter is used to maximise the margin between data points to classify the training sets properly. The testing dataset is represented into the same space to analyse the pollutant level.

33.4 Experimental Analysis

33.4.1 Study Area

This study estimates the air pollution level for Chennai City, capital of Tamil Nadu. Air pollution is found to be degenerating and life-threatening. The school children are suffering from irreversible lung damage. The residents are not aware of the effect of the air pollution, though they are affected a lot. Hence this study focuses on assessing the status of air quality data with respect to its effect on human health. Data has been collected from three major places in Chennai City – Alandur, Manali and Velachery. One-year data is being taken, and this dataset consists of parameters such as sampling date, state, city, location of monitoring station and the concentrations of air pollutants. The following pollutants, namely, CO, NO₂, PM_{2.5}, SO₂ and O₃, have been measured for the purpose of air quality status estimation. The dataset is divided into two sets: 90% for training and 10% of data are used for testing. The input for this model happened to be automatic monitoring of environmental data. The daily data are on the average of 24-h concentration value (8 hourly in case of CO and O₃) every month during the period of 1 year in 2017. The data are taken from the Central Pollution Control Board (CPCB), Ministry of Environment, Forests and Climate Change and Government of India (Website: www.cpcb.nic.in) which pinpoints the present status of air quality to control and normalise pollution.

33.4.2 Performance Metrics

This section focuses on the significant measures for judging this air quality assessment model. TP is the accurately categorised air quality status for each class. TN is the accurately uncategorised air quality status for each class. FP is the wrongly classified status of air quality for each class. FN is the wrongly unclassified status of air quality for each class.

The ratio of accurately categorised air quality level to the total dataset by the classifier gives the accuracy and is computed using Eq. 33.2.

$$\text{Accuracy} = \frac{TP + TN}{\text{dataset}} \quad (33.2)$$

The performance for the evaluation of status of air quality in Chennai City is given in Table 33.3. Comparative analysis of the classifiers is represented graphically in Fig. 33.2. It indicates that K-Nearest Neighbours provides 28.3% better accuracy than Random Forest and 52.83% better when related to Support Vector Machine.

Table 33.3 Performance metrics of the model using three different classifiers

Performance Metrics	K-Nearest Neighbour	Random Forest	Support Vector Machine
Accuracy	86.79	58.49	33.96
Precision	94.29	20.61	5.66
Recall	84.07	32.14	16.67
F1score	86.28	24.86	8.45

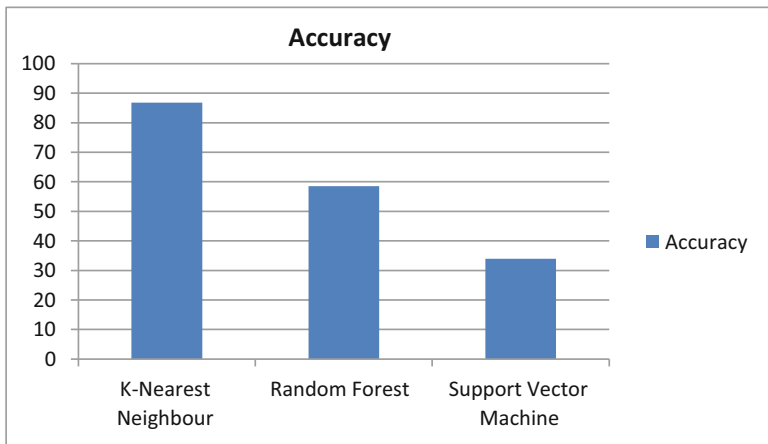


Fig. 33.2 Comparative analysis of the classifiers

The ratio of properly classified status of air quality to the total classified air quality by the classifier for that class offers the precision and is found using the Eq. 33.3.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{33.3}$$

The ratio of correctly classified air quality level by the classifier to all the observations in that class is represented as recall and is developed by Eq. 33.4.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{33.4}$$

Harmonic mean of Precision and Recall is the F1Score and is computed using Eq. 33.5.

$$\text{F1Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{33.5}$$

Figure 33.3 signifies F1 score Metric for the estimation of air pollution level using three classifiers, namely, K-Nearest Neighbour, Random Forest and Support Vector

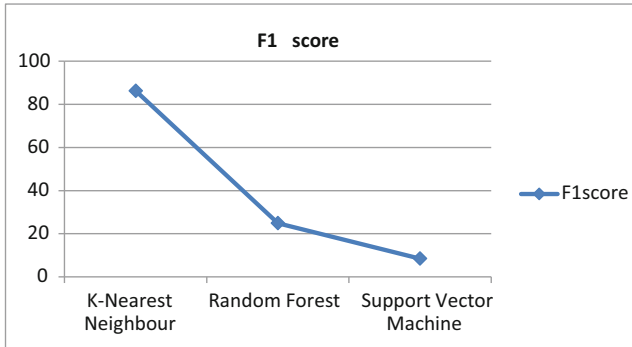


Fig. 33.3 F1score metric of the model using three different classifiers

Machine. It is found that K-NN provides the top performance in the classification of air quality levels compared to the others.

33.5 Conclusion

The Indian Government has taken numerous steps to mitigate air pollution. Since the AQI is represented in simple colours, any layperson can easily understand. A combination of two or more pollutants may be responsible for the pollution in a day, and so a common man ought to be aware of those pollutants to which they are sensitive.

This study clearly demonstrates the air pollution in India which rises every day results in poor ambient air quality. The quality of air is shown as Good, Satisfactory, Moderately Polluted, Poor, Very Poor and Severe so that the public can comprehend easily. The results clearly indicate that K-Nearest Neighbour-based estimation classifies the air quality significantly better than Random Forest and Support Vector Machine. The model will be a benchmark to the status of air inhalation not only in Chennai but also in other cities. Because of the maximum effort, study consequences and the mitigation measures are taken to the optimum level. This study renders help to the local residents of Chennai City so that they can plan their outdoor activities at the time of better air quality.

References

1. Vong C-M, Ip W-F, Wong P-k, Yang J-y (2012) Short-term prediction of air pollution in Macau using support vector machines. *J Cont Sci Eng* 2012:518032. 11 pages
2. Zhang C, Yuan D (2015) Fast fine-grained air quality index level prediction using random forest algorithm on cluster computing of spark. 978-1-4673-7211-4/15 IEEE

3. Dragomir EG (2010) Air quality index prediction using K-nearest neighbor technique. *Buletinul Universității Petrol – Gaze din Ploiești LXII(1):103–108*
4. Wang H, Jiao M, Tan Y (2016) Air quality index forecast based on fuzzy time series models. *J Res Sci Technol 13(5)*
5. Wang L, Bai YP (2014) Research on prediction of air quality index based on NARX and SVM. *Appl Mech Mat 602–605:3580–3584*
6. Mamta P, Bassin (2010) Analysis of ambient air quality using air quality index – a case study international *J Adv Eng Technol. E-ISSN 0976-3945.*
7. Zhao Y, Hasan YA (2013) Comparison of three classification algorithms for predicting pm 2.5 in hong kong rural area. *J Asian Sci Res 3(7):715–728*

Chapter 34

Implicit Continuous User Authentication Using Swipe Actions on Mobile Touch Screen with ANN Classifier



Christy James Jose and M. S. Rajasree

Abstract Smart phones became the most trusted companion of men and women. Today a smart phone has all the important data within the phone memory or in the cloud which is directly accessible by the device. The sensitivity of these data varies from person to person. Commonly, the security of a smart phone lies with normal entry point authentication methods such as PIN and graphical passwords. These methods can be breached by shoulder surfing, smudge attack, etc. But the main point is that the device is not checking the genuineness of the user after the entry point authentication. Most users prefer simple GUI passwords or PIN or no password at all (Bonneau J, The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: 2012 IEEE symposium on security and privacy, San Francisco, CA, pp 538–552, 2012). So a smart phone after the primary authentication can become a threat to losing sensitive and private data. In this work, we present an implicit continuous active authentication mechanism that will check the genuineness of the current user without any direct input. We would be using the touch screen swipe patterns that are being generated when reading a page or viewing the images in gallery of the phone. We are using the artificial neural network to recognize the genuine user. Result shows the proposed mechanism has an accuracy of 93.9% and an EER of 7%, and it is not a burden to the user as he is not supposed to make any deliberate inputs, the data generated from the normal usage is taken for authentication.

Keywords Implicit authentication · Information security · Smart phone · Swipe biometrics

C. J. Jose (✉) · M. S. Rajasree
Government Engineering College, Thiruvananthapuram, Kerala, India

© Springer Nature Switzerland AG 2020
L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_34

Abbreviations

ANN	Artificial Neural Network
PIN	Personal Identification Number
SVM	Support Vector Machine
KNN	K Nearest Neighbor
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
FAR	False Acceptance Rate
FRR	False Rejection Rate
EER	Equal Error Rate
csv	Comma Separated Value
GUI	Graphical User Interface

34.1 Introduction

Smart phones are getting smarter day by day, and the number of users enjoying it to ease their daily life is increasing. These categories of people are generating and saving more and more sensitive and private data. In India the number of digital transactions in 2017 had marked 32% increase and a 42% increase in mobile banking amounts to that of 2014 [2]. The objective of an attacker would be to gain access to the victim's phone so that he can steal or corrupt the private data on it. There are two approaches for authentication in smart phones which uses the touch screen. One is the Entry Point authentication or the primary authentication via the secret PIN or Graphical Pattern. It is found that biometric features like the gait, face, keystroke, voice, touch gestures, and swipe patterns are generating sensor data in the background. The same could be used to verify the genuineness of the user. Based on this, another method known as continuous authentication has emerged. This is also known as Implicit Authentication or Transparent Authentication. Here the sensor data generated while the user interacts with the phone for his routine activities are recorded in the background without interference to the user, and these data are used to verify the genuineness of the current user with the help of a suitable classifier [3–5]. A continuous authentication would be appreciable by the user if and only if the mechanism would put little or no additional burden on him since even the entry point authentication itself is a burden for most of the users [1]. In the described scenario, we suggest an implicit mechanism which relies on the basic and mostly used action on the touch screen of the mobile device that is the swipe actions performed for viewing the gallery images and reading a document. In this paper, we are

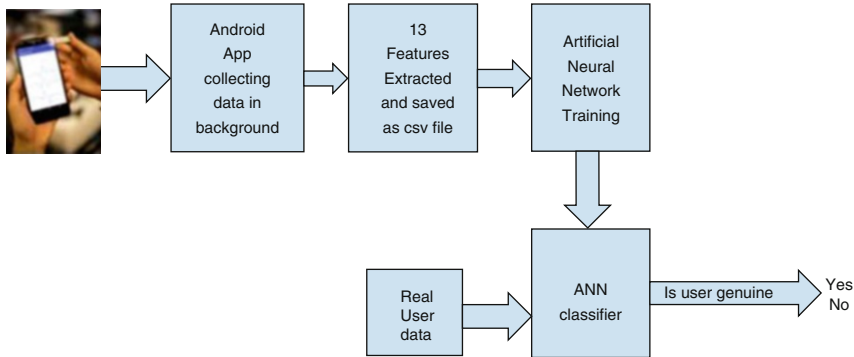


Fig. 34.1 Block diagram of the proposed work

investigating the possibility of using the swipe actions on a mobile touch screen to authenticate the genuine owner of the device. We have used the Artificial Neural Network to classify the genuine user. Our proposed methodology is not demanding any exclusive input from the user, rather the data from the sensors are collected in the background silently and are used for computation. Results show that the swipe data when used with an artificial neural network on supervised learning gives promising results to use them for continuous authentication of the user in real time. Our main goal is to deny access to the attacker without creating any inconvenience to the genuine user. Figure 34.1 shows the block diagram of our proposed work.

34.2 Related Works

Biometrics are essentially human characteristics which refer to the body movements and calculations. Biometric authentication has a long history. In the case of mobile devices, it was limited to fingerprint sensors and face unlocks which fall under the entry point authentication schemes. Jakobsson et al. [3] analyzed the possibility of authenticating the genuine user implicitly by making use of multiple data. They have used the location data, information about the network connections, application installation and usage, call logs, SMS, and calendar entries. Their algorithm computes a score based on the above features. Access to the device is based on the comparison of the current score computed to the score computed during training for the genuine user. Score could be seen as the probability of the device being used by the genuine user. Keystroke dynamics-based authentication which was originally envisaged for desktop computers which uses a hardware keyboard was extended to mobile phone touch screens where a virtual keyboard is used. In this type of

authentication scheme, the sensor data along with the typing behavior is recorded and used to make a distinction between the genuine user and intruder [4]. The following features are considered: time taken in between the pressing of two keys, usual delay in between key strokes, time taken for a single key press and release, and the number of times the backspace key is pressed. The above features were used to classify the genuine user and an intruder. Most of the works in this method used traditional classifiers like Naive Bayes, Hamner distance, Euclidean distance, etc. Neural Networks were used in [5] to have an adaptation and better results. Gait dynamics-based authentication schemes rely on the walking pattern of an individual. Here the data collection is to be done using a wearable sensor. Statistical classifiers like Support Vector Machine (SVM) were used in the referred works [10]. In the Touch Dynamics-based authentication, the way the user touches his device screen for swipes, gestures, etc., are considered [8, 9]. Our work lies in this category. During the touch actions, the features like size of the finger in contact with the screen, the pressure by finger on the screen, x- and y-coordinates, accelerometer readings, time stamp, etc., could be recorded. These extracted features can be used for continuous user authentication. In [6], the authors verified the possibility of using the touch data to classify the genuine user in continuous mode with a database of 41 users with a 27-dimensional feature vector with the help of Support Vector Machine and K Nearest Neighbor classifiers. Some other works verified the single touch operation for authentication [7, 10]. They have mentioned that SVM with RBF Kernel, Random Forest, and Logistic regression were better with 10–20% EER. In [10], they have used combination of swiping, typing, and movement of phone. All the works had accuracy above 90%.

34.3 Materials and Methods

Swipe action is inevitable in a smart phone having touch screen. Any user whether genuine user or attacker needs to use minimum number of swipes on the touch screen to access the services offered by the device. For the experiment, we need a data set. In this work, we have generated our own data set. An android application was developed to record the sensor data that is being generated during the swipe actions. The app when started will record all the 13 features associated with a swipe action whenever a user is viewing the gallery or reading a document.

The recording is done in the background silently without any burden on the user. The app will save each and every session as a *.csv file. The data set consists of swipe data of ten individuals. All of them were post-graduate students of our institute: four males and six females whose ages were in the range of 24–26 years. The data generation was done in a single session. All subjects were asked to view the image gallery and to read a document in document reader. This activity was carried out after initializing the specially built android app. In our work, we are considering

X value	Y value	Z value	Size	raw x	raw y	touchmajor	touchminor	x	y	xvelocity	yvelocity	distance	Label
-0	9.03	3.64	0.27	484	863	4	4	452	671	-441	-232	16.2	0
0.01	9.58	1.93	0.2	630	867	3	3	598	675	-324	0.02	13.4	0
1.79	5.66	7.78	0.13	566	790	2	2	534	598	-3395	937	266	0
1.75	5.65	7.98	0.27	522	751	4	4	490	559	-5853	2751	219	0
1.54	5.62	7.89	0.2	286	822	3	3	254	630	-0.03	-0.1	0	0
-0.3	-0.1	9.79	0.07	546	714	1	1	514	522	1658	-387	67.3	1
-0.3	-0.1	9.79	0.07	544	703	1	1	512	511	675.6	-74	20	1
-0.3	-0.1	9.79	0.13	593	688	2	2	561	496	716	-96	18.2	1
-0.3	-0.1	9.79	0.27	569	690	4	4	537	498	1260	-29	45.1	1
-0.3	-0.1	9.79	0.2	570	701	3	3	538	509	1981	-293	78.9	1
-0.5	4.27	8.71	0.13	617	732	2	2	585	540	0.004	-71	3.13	1

Fig. 34.2 Sample of data set

the action by thumb finger only, and all the subjects were directed to use the portrait mode. All were using the same device for data collection, in this case a Samsung J5 phone. This is to ensure the data set uniformity as the device parameters like screen size and the sensor availability in different devices might affect the setup. All the subjects were directed to use the thumb finger only to swipe, and all were asked to sit in a chair during the activity. We thought the usage of thumb finger will be easiest. It has to be performed with almost all applications in the phone. Most importantly, it does not pose any burden on the user. Out of the ten individuals, the data of one user was selected as the genuine user and all others as impostors. So we have recorded more sessions of activity of the chosen genuine user. In total, we have 145 samples with 13 features from 10 individuals. Out of which, 91 samples were of the genuine user. A sample of the collected data is given in Fig. 34.2.

34.4 Feature Explanation

34.4.1 x value, y value, and z value

This feature gives the orientation of smart phone. The way in which one holds the smart phone is entirely different from another. These are the accelerometer readings in x, y, and z axes plotted in Fig. 34.3. The label 1 is for the genuine user samples and 0 denotes others. We can see that this data has enough variance to distinguish between the genuine user and an impostor, and it supports our statement that everyone holds a phone in a unique manner.

X value, Y value, Z value and Label

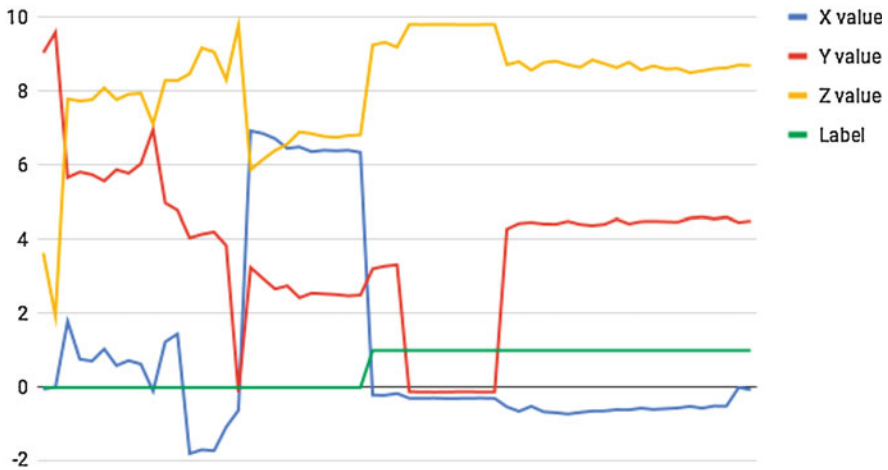


Fig. 34.3 Accelerometer readings

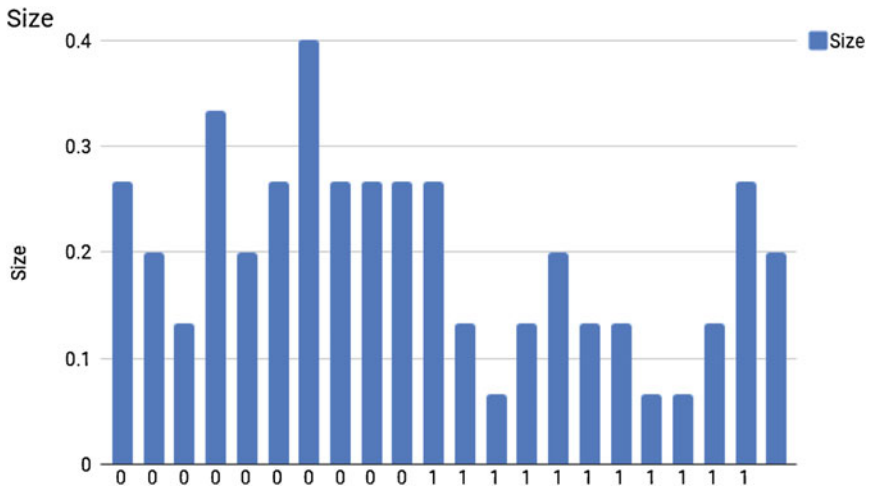


Fig. 34.4 Finger size

34.4.2 Size

This feature is the finger size, and it represents the area covered by the finger on the touch screen. This is illustrated in Fig. 34.4. Finger size varies among different age categories and gender to some extent. This data too shows enough distinction between values even though some of the genuine user samples are not uniform.

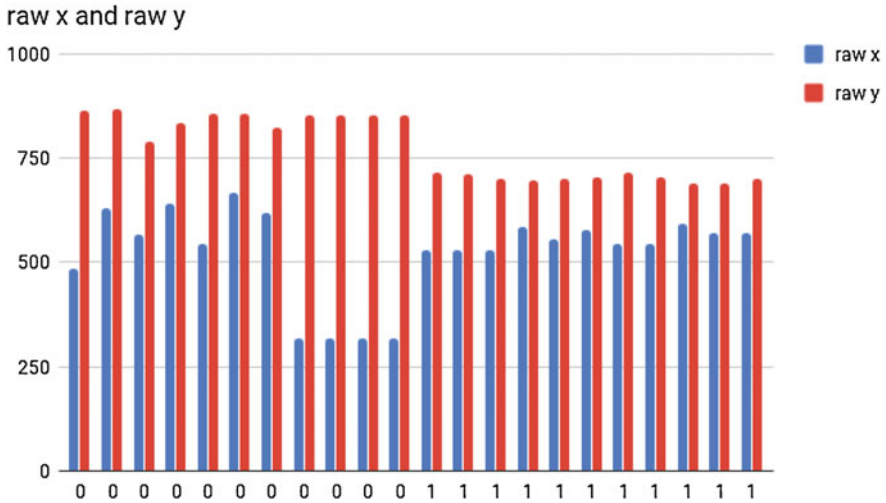


Fig. 34.5 Relative coordinates

This is because the finger area that makes contact with the touch screen is not the same always.

34.4.3 Raw x, raw y

These are the relative coordinates. They are coordinates related to a view, and their origin is not fixed. These parameters can be used to represent the behavior of a touch screen user. Some users prefer the right side of the touch screen while others the left side. Coordinates indicate the location preference of each user. This is illustrated in Fig. 34.5.

34.4.4 Touch Major, Touch Minor

This gives the pressure on touch screen, and it varies from user to user based on how hard they exert force on the touch screen (Fig. 34.6).

34.4.5 x, y

The absolute coordinates represent coordinates of the touch screen, and their origin is fixed. This gives the location preference of each user (Fig. 34.7).

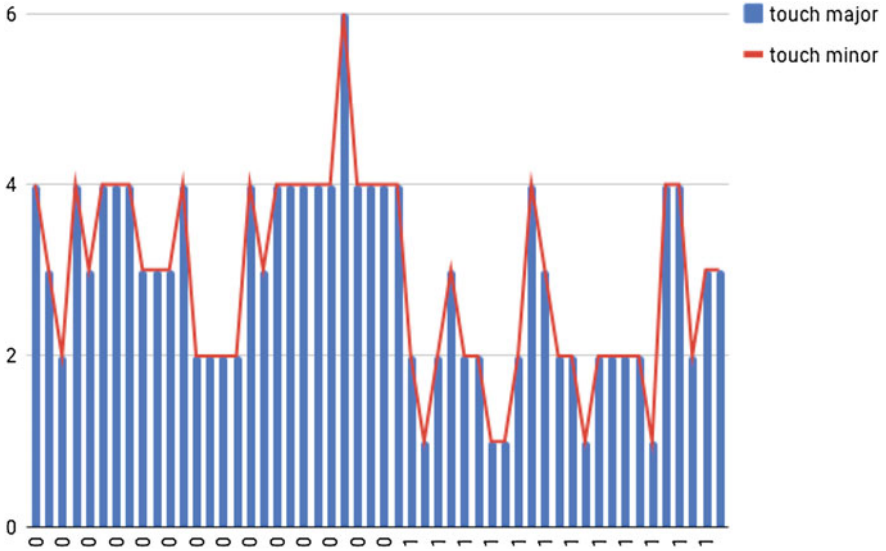


Fig. 34.6 Pressure on touch screen

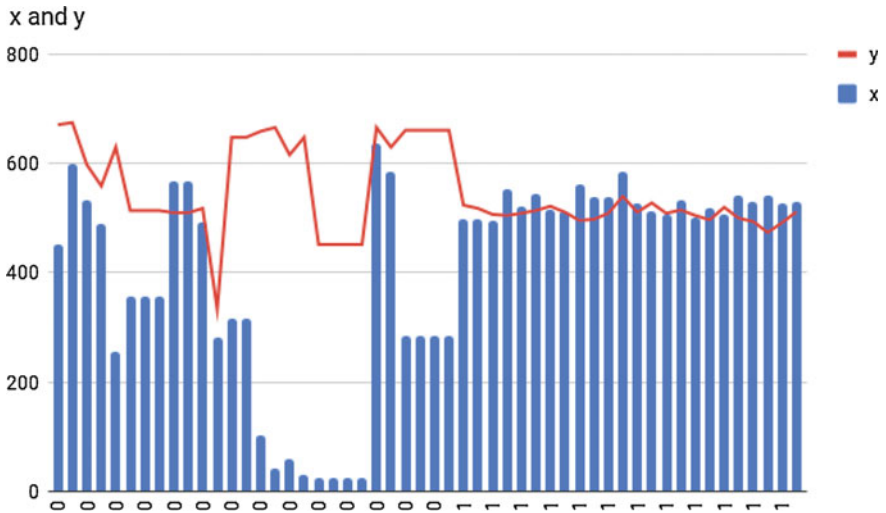


Fig. 34.7 Absolute coordinates

34.4.6 *x velocity, y velocity, and Distance*

Speed or velocity of swiping varies with time and distance covered during swiping. Some users cover a particular screen area in a short duration of time, whereas some others take long duration. Velocity component along x and y directions is taken.

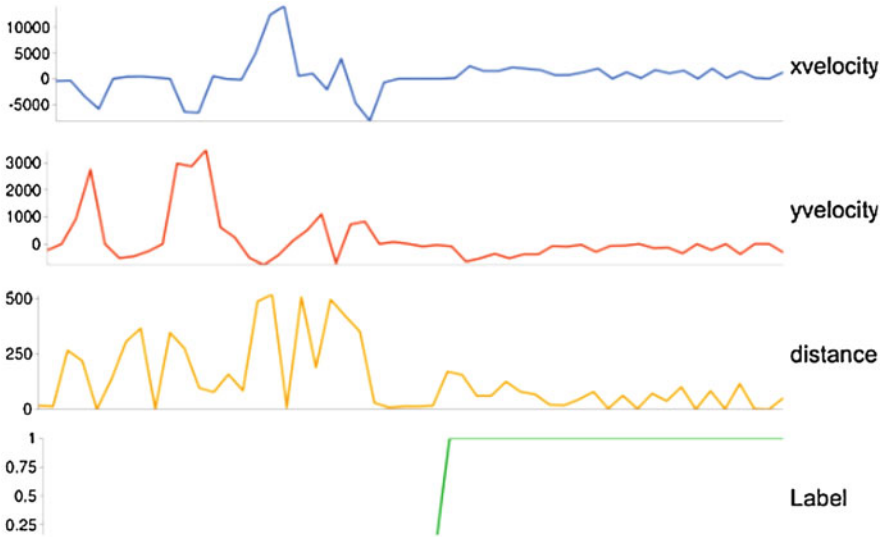


Fig. 34.8 Velocity and distance

Distance gives the length of the swiped line. This feature too gives unique patterns of swiping movements. Label 1 is for the genuine user. All other samples are labeled 0 and considered impostor (Fig. 34.8).

34.5 Results and Discussion

Most of the related works relies on the popular classifiers like Support Vector Machine (SVM), K Nearest Neighbor (KNN), etc. Some of the works mentions fusion of swipe along with some other data such as keystroke and movement patterns. In our work, we are proposing artificial neural networks to recognize the genuine user. Neural networks perform in an excellent way compared to other classifiers if there are a lot of data available and the structure is complex. They can have many layers.

The disadvantage is that we need a lot of samples per user. Ours is a binary classification problem. Our input data set is associated with different classes, ANN can classify the known input properly, and it will accurately classify the inputs that were not part of the solution design. The 13 attributes listed above are given as input training data, and the respective targets are 0 and 1. The attributes corresponding to genuine users are mapped to 1 and others to 0. We have used a Multilayer Perceptron which comes under the category of Supervised Learning algorithm. It can learn a function by training a data set. If a set of features and a target is available, it can learn a nonlinear function approximator for either classification or regression. This is one

Table 34.1 Performance of the system

Sl No	All measures in percentage		
	Measure	Value	Derivation
1	Accuracy	93.79	$ACC = (TP + TN) / (P + N)$
2	Specificity	92	$SPC = TN / (FP + TN)$
3	Sensitivity	97.83	$TPR = TP / (TP + FN)$
4	EER	7	Crossing Point FAR and FRR
5	False Positive	8.08	$FPR = FP / (FP+TN)$

of the most widely used neural network classifiers. It is competent in modeling complex functions, robust and adaptive to weights and environment changes. Number of output is 1 as either genuine user or not. All the simulation was carried out using the Neuroph Studio. We have used one hidden layer and 14 hidden neurons. We have used Back propagation with momentum as learning rule. Figure shows the Neural Network architecture. Results show our ANN design is efficient in classifying the patterns.

Being a binary classification problem, we can determine its performance by working out the number of genuine users labeled correctly (True Positive), the number of imposters labeled correctly (True Negative), number of imposters labeled as genuine (False Positive), and number of genuine users classified as imposters (False Negative). These four values will be used to constitute a Confusion Matrix. Another metric used is the Accuracy which is the ratio of Number of Correct Predictions to the total number of Predictions. Sensitivity is the ratio of correctly classified positive observations to the actual positive observations. Specificity is the ratio of correctly labeled negative samples to the actual number of negative samples. It is the efficiency of classifier in identifying negative samples. Equal Error Rate (EER) equates to the point at which FAR and FRR cross. Lower the EER higher the accuracy of the system. Table 34.1 describes the performance of the system.

34.6 Conclusion

We have investigated the possibility of using simple touch movements, which are part of activities like viewing the images in gallery or reading a document or web page on a smart phone touch screen for continuous authentication. The results show they are sufficient for continuously authenticating the genuine user. The limitations of this work are described below. It was seen that not all the available smart phones have the ability to extract all the 13 features studied in this work. The feature that would be missing in some of the android phones would be pressure and finger size. So as a future work, we can check whether in the absence of these features, the system would work efficiently. Or will it change the performance figures? Moreover our data set was generated in a controlled environment. Same task was given to all subjects, and all subjects were in a sitting position. They were directed to use their

thumb finger only in portrait mode only. This may not model a real-time usage pattern. In normal usage, everyone has their own preferences of fingers; landscape mode might be chosen over portrait and not necessarily be sitting all the time. Also, we have not studied the effect of different screen size, effect of user movements, orientation of phone, and the swipe using fingers other than the thumb finger. So we have to modify the data collection application to record the real-time usage in the background. Compared with existing classifiers, Artificial Neural Network Classifier gives better accuracy and EER.

References

1. Bonneau J (2012) The science of guessing: analyzing an anonymized corpus of 70 million passwords. In: 2012 IEEE symposium on security and privacy, San Francisco, CA, pp 538–552
2. Dutta S, Edited by Sonia Bhaskar, Increase in digital transactions in a year; mobile banking jumps 122%. Available at <https://special.ndtv.com/cashless-bano-india-14/news-detail/india-sees-55-increase-in-digital-transactions-in-a-year-mobile-banking-jumps-122-1724624/7>
3. Jakobsson M, Shi E, Golle P, Chow R (2009) Implicit authentication for mobile devices. In: Proceedings of the 4th USENIX conference on Hot topics in security, pp 9–19
4. Crawford (2010) “Keystroke dynamics: characteristics and opportunities” Privacy Security and Trust (PST). In: 2010 eighth annual international conference, pp 205–212
5. Jose CJ, Francis J, Rajasree MS (2015) Digraph approximation with an adaptation technique for mobile user authentication through keystroke dynamics. *Int J Eng Adv Technol* 5(1). ISSN: 2249–8958
6. Derawi FM, Bours P (2013) Gait and activity recognition using commercial phones. *Comp Sec* 39:137–144
7. Feng et al Continuous mobile authentication using touchscreen gestures. In: 2012 IEEE Conference on Technologies for Homeland Security (HST), Waltham, MA, vol 2012, pp 451–456
8. Xu H, Zhou Y, Lyu MR (2014) Towards continuous and passive authentication via touch biometrics: an experimental study on smartphones. In: Proceedings of SOUPS, pp 187–198
9. Shen C, Zhang Y, Guan X, Macion RA (2016) Performance analysis of touch-interaction behavior for active smartphone authentication. *IEEE Trans Info For Sec* 11(3):498–513
10. Kumar R, Phoha VV, Serwadda A (2016) Continuous authentication of smartphone users by fusing typing swiping and phone movement patterns. In: Proceedings of IEEE BTAS, pp 1–8

Chapter 35

A Review on Graph Analytics-Based Approaches in Protein-Protein Interaction Network



D. Narmadha, A. Pravin, G. Naveen Sundar, and Premnath Dhanaraj

Abstract Essential proteins play a vital role in the biological and cellular activity of a living organism. Identification of essential proteins is crucial for understanding the cellular life mechanisms for medical treatments and disease diagnosis. The existing computational measures are primarily based on identifying dense sub-graphs from the protein interaction network. In this research paper, the existing computational, graph theoretic approaches are reviewed and a novel research direction to find essential proteins is proposed.

Keywords Protein-protein interaction · Centrality measure · Graph theory · Essential proteins · Drug discovery · Computational methods · Knowledge discovery · Unsupervised and supervised methodologies

Abbreviations

PPI	Protein-protein interaction
RNA	Ribonucleic acid
Bio GRID	The Biological General Repository for Interaction Datasets
PPIM	Protein-protein interaction database for maize
DIP	Database of Interacting Proteins

D. Narmadha (✉)

Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

e-mail: narmadha@karunya.edu

A. Pravin

Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

e-mail: pravin.cse@sathyabama.ac.in

G. Naveen Sundar · Premnath Dhanaraj

Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

e-mail: naveensundar@karunya.edu; premnath@karunya.edu

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_35

SGD The *Saccharomyces Genome Database*
MIPS Munich Information Center for Protein Sequences

35.1 Introduction

The existence of living organisms is dependent on the cellular level activities proceeding inside its body. The most important cell activities such as respiration, muscle contraction, and protein synthesis from amino acids are controlled by the protein network of the organism. In the protein network, all the proteins are not equally the same. Essential proteins are crucial for the sustainability of living organisms [1]. Studies have also shown that deletion of an essential protein can cause the lethality of the cells [2]. Essential proteins are extremely important for developing novel drugs and thereby help to protect someone from deadly diseases [3, 4]. Experiential methods such as conditional knockouts [5], single-gene knockouts [6], and RNA interference [7] are quite complex, consume more time, and are not suitable for all the organisms. These methods also do not suit all the organisms. To overcome the challenges of these approaches, computational methods have been introduced to work with genome data, protein sequence data, protein-protein interaction data, and cDNA and gene expression data.

The computational methods [8] are generally categorized into local and global topological properties. Besides the vast number of experimental methods, a huge collection of biological databases are available which contain the information about PPI for a variety of species. Some of the most useful biological databases are IntAct [9], BioGRID database [10], Molecular Interaction database [11], Protein-Protein interaction database for Maize [12], STRING database [13], BioCreative database [14], DIP database [15], SGD database [16], and MIPS database [17].

The purpose of this literature survey is to provide insight into the latest development in the field. The organization of the paper is as follows. In the second section, computational methods such as graph theoretic measures and unsupervised and supervised learning are discussed. In the third section, the pitfalls and advantages of various computational methods are discussed. Finally, a future protocol for research problem is discussed.

35.2 Essential Protein Identification

The identification of essential proteins can be done from 2D sequence of amino acids, protein interaction data, and 3D protein structure data. Rule-based induction decision trees, functional trees, and instance-based learning are used to find information about individual proteins in 2D protein sequence data. Computational approaches such as graph-theoretic measures and unsupervised and supervised machine learning methods are used to find the essential proteins in the 3D protein structure.

35.3 Computational Methods

Proteins are the most important building blocks of cells. Proteins are molecules made up of several chains of amino acids in a definite order. Each protein is essential for the functioning and regulation of body organs and tissues. The protein does not act alone. It interacts with other proteins to perform biological function [18]. Several types of research have been carried out in the recent past to explore the relationship between essentiality of protein with the graph theoretic measures. Graph theoretic measure is generally classified into local and global properties. The measure that is applied only to the small sub-graph is known as local properties. The measure which is applied globally to the entire network is known as global property.

To provide a simplified description of the usage of graph theoretic measure for essential protein identification, there are different ways of representing graphs mathematically such as directed, undirected, weighted, and bipartite graphs.

Figure 35.1 shows the directed graph representation of nodes and edges. In this figure, $V = \{N, O, R, Q\}$ represents the set of vertices and $E = \{(O, N), (N, R), (Q, N), (N, Q), (O, Q)\}$ represents sets of edges and $|E| = 5$ which represents the sets of edges.

Figure 35.2 shows the pictorial representation of undirected graph where $V = \{I, J, K, L\}$ represents the set of vertices and $E = \{(I, J), (I, K), (I, L), (J, L)\}$ represents the set of edges and $|E| = 5$ which represents the number of edges.

Figure 35.3 shows the pictorial representation of weighted graph $V = \{A, B, C, D\}$ that represents the set of vertices and $E = \{(D, A, 3), (A, B, 7), (C, B, 5), (D, B, 11)\}$ $|E| = 4$ which represents the number of edges.

Fig. 35.1 Directed graph

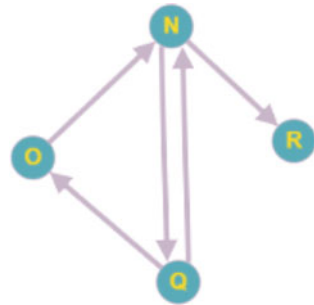


Fig. 35.2 Undirected graph

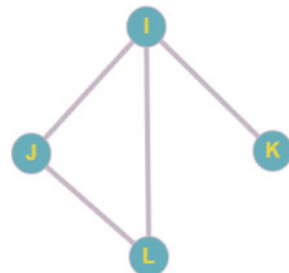
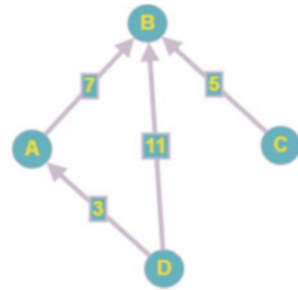


Fig. 35.3 Weighted graph



The following section describes the various classifications of topological features.

35.3.1 Degree Centrality

Degree centrality (DC) [19] is a category of graph theoretic measure that helps to detect essential nodes that affect the topology of the network. It is one of the most important graph theory measures that calculate the number of incoming edges and outgoing of a node. For a node x , the degree centrality of x is calculated using the following equations. The topology of the biological network is greatly affected when such prominent nodes of high degrees are removed. The research idea which has been discussed in [20] throws light on the importance of finding degree centrality that paves way for prediction of pathway analysis and protein functionality.

$$Deg_x = Deg_{in}(x) \quad (35.1)$$

$$Deg(y) = Deg_{out}(y) \quad (35.2)$$

The number of incoming edges that is incident on a particular node is calculated using Eq. 35.1 and the number of outgoing edges from a particular node is calculated using Eq. 35.2.

35.3.2 Betweenness Centrality

Betweenness centrality (BC) [21] is used to measure the most targeted node in a network. It is most widely used for analyzing social network and protein interaction network. It is the measure of the percentage of the shortest path that passes through a node. Betweenness centrality of a node Y is designed using the subsequent formulae.

$$BC(Z) = \sigma_{AB}(Z) / \sigma_{AB} \quad (35.3)$$

where $\sigma_{AB}(Z)$ denotes the number of small paths between nodes A and B that passes through Z . The node with high betweenness score tends to be the most essential

nodes than all other nodes [22]. Nodes with high betweenness centrality drive the signaling pathway for diseases and the importance of betweenness centrality is discovered in the yeast network [23].

35.3.3 Closeness Centrality

Closeness centrality (CC) of a node x is the reciprocal of the shortest path distance from the node x to all other nodes in the network. It is the only measure in which a lower value is better. The nodes with the lower measure tend to be closer to other nodes in the network. Closeness centrality of a node is calculated using the following formula. In the research insights provided by Girvan M, Newman ME [24], and Hahn MW, Kern AD [25] also portrays the importance of applying closeness centrality measure in different types of the protein interaction network.

$$CC(X) = \frac{\text{Total of the shortest distance from } x \text{ to other nodes}}{\text{Total no.of nodes}} \quad (35.4)$$

35.3.4 Eigen Vector Centrality

Eigen vector centrality helps to compute node's influence in a network. It uses iterative matrix computation to assign scores to the nodes in the network. More score is given to the nodes if they are connected to influential nodes. [26] shows the application of eigen vector centrality in extracting a research professional's relationship. [27] shows the importance of using eigen vector centrality to explore the disease-gene relationship.

35.3.5 Modularity and Community Structure in Networks

Modularity and Community (MC) [28] focuses on the problem of detecting the community structures. The network is divided into communities. Modularity matrix (eigen values) is used to analyze the community structure. It gives more weight to the node if they are connected to influential nodes. This approach consumes more time to calculate the eigen vector of the modularity matrix. Bennett et al. [29] discusses the importance of detecting overlapping communities using a novel mathematical function which finds its application in yeast and human organisms. [30] explores the importance of extracting homologous communities within a network.

35.3.6 Centrality and Modularity

Centrality and Modularity (CM) uses graph theoretic measures such as centrality and modularity to determine the protein pathway. First-order and second-order interactions are obtained from the STRING Database. The analysis is done on both interactions to determine the proteins which are responsible for viral infections. Vallabhajosyula RR analyzed the importance of identifying hub nodes and its correlation to the extraction of the biologically significant proteins. Song J and Singh M infer that there is a very strong relationship between hub proteins and essential proteins (Table 35.1).

Table 35.1 Analysis of computation-based approaches

S. No.	Database	Methods	Input	Functionality	Accuracy
1	StringDB	Degree centrality	Structure	Computes the node that affects the network topology	0.72
2	StringDB	Betweenness centrality	Structure	Assesses the important node by computing the percentage of the shortest path	0.682
3	DIP	Closeness centrality	Structure	Determines how close a node is from the target node	0.687
4	DIP	Eigen vector centrality	Structure	Iterative matrix computation to measure nodes importance	0.722
5	Metabolic network	Modularity and community structure in Network	Structure	Eigen vectors and spectral clustering are used for community detection	0.654
6	SGD, DEG	Centrality and Modularity	Structure	Proteins are ranked in terms of connectivity and degree of the nodes	0.627
7	MIPS	Essentiality and Centrality	Structure	Proteins that are reached by single interactions are given higher ranks	0.678
8	MIPS, SGD, DEG	United Complex Centrality (UCC)	Structure	Topological and protein structural data integration	0.687
9	MIPS, SGD, DEG	LBCC-based approach	Structure	Proteins are ranked based on combination of centrality measures	0.784
10	DIP	Features with respect to direct interaction	Structure	Ranking is done by assigning more weights to the proteins that have many interacting neighbors	0.77
11	DIP	Prior knowledge-based method to find essential proteins	Structure	High ranks are assigned to the proteins that have high probability of becoming the known essential proteins	0.79

35.4 Protocol

In the research review, we have reviewed all existing computation-based approaches and unsupervised approaches to predict the essential proteins from the protein interaction network. The existing approaches work based on choosing a random molecule in a network and then proceed with a walk-through along the networks to predict the other essential proteins. Here, a new insight into the task of finding the essential proteins from the protein interaction network is proposed. Graph coloring-based approach has been proposed to find the target and the essential proteins from the network. Graph coloring is a very imperative methodology which is used most widely in many areas of computer science such as protecting art gallery by installing cameras, scheduling timetable for examination, scheduling aircrafts to flights, coloring the maps, and identifying signaling pathways for diseases. However, the graph coloring algorithm does not find any application to extract the essential proteins from the biological protein interaction network.

35.5 Conclusion

Multiple mathematical models are available to analyze PPI networks, but it is more daunting to understand the whole bio mechanism. But multidisciplinary approaches need to be considered for big data analysis. Even though existing algorithms provide good result, novel methods are more useful to improve the perfectness and its ability to scale for larger datasets. The usage of the mathematical models such as graph-theoretic measures and clustering produces some insight information about the protein complexes. However, there are still open challenges in detecting protein complexes. The existence of proteomes has been hypothetically analyzed, and the proposed graph coloring algorithm results are used to predict PPI through protein-conserved relationships. Here, we are approaching the moment when enough information would be available to verify the existence of such proteome and discover its structural properties within PPI networks. Complex biological networks show that the usage of graph theoretic properties reflects information to understand the biological mechanisms. Similar analysis has been applied to analyze PPI by integrating software to view the importance of node in graphical mode. In the modern days using interdisciplinary fields such as biology and information technology helps to analyze highly complex network in a short span of time. Hence novel algorithms are needed to analyze complex networks. In view of this review paper, a novel method of coloring algorithms to predict PPI and centrality protein networks is proposed.

References

1. Pál C, Papp B, Hurst LD (2003) Genomic function (communication arising): rate of evolution and gene dispensability. *Nature* 421:496
2. He X, Zhang J (2006) Why do hubs tend to be essential in protein networks? *PLoS Gen* 2:e88
3. Fuentes G et al (2011) Role of protein flexibility in the discovery of new drugs. *Drug Dev Res* 72:26–35
4. Clatworthy AE, Pierson E, Hung DT (2007) Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol*. 3:541
5. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, Tandia F, Linteau A, Sillaots S, Marta C (2003) Large-scale essential gene identification in candida albicans and applications to antifungal drug discovery. *Mol Microbiol* 50:167–181
6. Xu Z, Zikos D, Osterrieder N, KarstenTischer B (2014) Generation of a complete single-gene knockout bacterial artificial chromosome library of cowpox virus and identification of its essential genes. *J Virol* 88:490–502
7. Walia RR, Caragea C, Lewis BA, Towfic F, Terribilini M, El-Manzalawy Y, Dobbs D, Honavar V (2012) Protein-RNA interface residue prediction using machine learning: an assessment of the state of the art. *BMC Bioinform* 13:89
8. Qin C, Sun Y, Dong Y (2016) A new method for identifying essential proteins based on network topology properties and protein complexes. *PLoS One* 11:e0161042
9. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A, Margalit H (2004) Int Act: an open source molecular interaction database. *Nuc Acid Res* 32(Suppl_1):D452–D455
10. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M (2006) Bio GRID: a general repository for interaction datasets. *Nuc Acid Res* 34(Database issue):D535–D539
11. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the molecular interaction database. *Nuc Acid Re* 35(Database issue):D572–D574
12. Zhu G, Wu A, Xu X-J, Xiao P-P, Lu L, Liu J, Zhao X-M (2016) PPIM: a protein-protein interaction database for maize. *Plant Physiol* 170:618–626
13. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, von Mering C (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucl Acid Res* 45(Database issue):D362–D368
14. Arighi CN, Roberts PM, Agarwal S, Bhattacharya S, Cesareni G, Chatr-aryamontri A, Wu CH (2011) Bio Creative III interactive task: an overview. *BMC Bioinform* 12(Suppl 8):S4
15. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D (2000) DIP: the database of interacting proteins. *Nuc Acid Res* 28:289–291
16. Luo H, Lin Y, Gao F, Zhang CT, Zhang R (2013) DEG 10, an update of the database of essential genes that includes both protein-coding genes and noncoding genomic elements. *Nuc Acid Res* 42:D574–D580
17. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW, Ruepp A (2004) The MIPS mammalian protein–protein-interaction database. *Bioinformatics* 21:832–834
18. Vazquez A, Alzate O (eds) (2010) Protein interaction networks, neuroproteomics. CRC Press/Taylor & Francis, Boca Raton
19. Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein–interaction networks. *Mol Biol Evol* 22:803–806
20. Mistry D, Wise RP, Dickerson JA (2017) Diff SLC: a graph centrality method to detect essential proteins of a protein–protein interaction network. *PLoS One* 12:e0187091
21. Opsahl T, Agneessens F, Skvoretz J (2010) Node centrality in weighted networks: generalizing degree and shortest paths. *Social Net* 32:245–251
22. Abedi M, Gheisari Y (2015) Nodes with high centrality in protein interaction networks are responsible for driving signaling pathways in diabetic nephropathy. *Peer J* 3:e1284

23. Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network. *Bio Med Res Int* 2:96–103
24. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Nat Acad Sci USA* 99:12–7821–7826
25. Hahn MW, Kern AD (2004) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 22:803–806
26. Bihari A, Pandia MK (2015) Eigenvector centrality and its application in research professionals' relationship network. In: *Futuristic trends on computational analysis and knowledge management (ABLAZE)*, pp 510–514
27. Özgür A, Vu T, Erkan G, Radev DR (2008) Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 24:i277–i285
28. Newman ME (2006) Modularity and community structure in networks. *Proceed Nat Acad Sci* 103:8577–8582
29. Bennett L, Kittas A, Liu S, Papageorgiou LG, Tsoka S (2014) Community structure detection for overlapping modules through mathematical programming in protein interaction networks. *PloS One* 20:e112821
30. Lewis AC, Jones NS, Porter MA, Deane CM (2010) The function of communities in protein interaction networks at multiple scales. *BMC Syst Biol* 4:100

Chapter 36

A Survey on Emotion Detection Using EEG Signals



Oshin R. Jacob and G. Naveen Sundar

Abstract Emotions play a huge role in the social interactions between people which makes it important to study its working to make intelligent humanoids that can socialize on a higher dimension than was deemed possible in the last century. This paper gives a survey on the various approaches used to detect emotions and explains the role of the brain in generating human emotions. A review is also made on the available classifiers and latest trends used in analyzing the EEG signals.

Keywords EEG · SVM · Valence arousal · Emotions · Physiological signals

Abbreviations

AI	Asymmetry Index
DEAP	Database for Emotion Analysis Using Physiological Signals
ECG	Electrocardiogram
EEG	Electroencephalogram
ERNN	Elman Recurrent Neural Network
GSR	Galvanic Skin Response
KNN	K-Nearest Neighbor
PPG	Photoplethysmography
PSD	Power Spectral Density
SVM	Support Vector Machine

O. R. Jacob (✉) · G. Naveen Sundar (✉)
Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India
e-mail: oshinrjacob@karunya.edu.in; naveensundar@karunya.edu

36.1 Introduction

The interest over the “seat of emotions” has been a long debated topic since the age of the philosophers. The fields of art and science have a conflict of interest even today, when it comes concluding the “seat of emotions” as the brain or heart. Medicine claims the amygdala of the brain to be the originator for emotions, while philosophers and poets emphasize the heart to be the carrier of emotions. In this age where the growth of artificial intelligence and the creation of humanoids are hitting the pinnacle, scientists, developers, and researchers are exploiting the human anatomy to make intelligent systems. The field of artificial intelligence embeds within it many systems that are designed based on the working of the human brain, like the neural networks, expert systems, machine learning algorithms, and deep learning techniques. This concept of mimicking the human brain is led by the popular notion that brain is the main organ that controls the human body. But recent studies have taken the best from the world of philosophers and medicine by making systems that detect emotions from physiological signals as well as outward expressions of the body. However, the techniques used vary in their accuracy and complexity, thus making it vital to select the best algorithms, features, and parameters to detect emotions. This paper attempts to find a good solution toward an optimized way to detect emotions.

36.2 Related Work

The research to manipulate emotions have been exploited widely to incorporate its working in the field of artificial intelligence. Emotional response can be extracted in two ways, that is, outward expression of emotions and the inner feelings. The outward expressions are captured from facial expressions, manner of speech, and other bodily movements. Popular existing humanoids that have incorporated these technologies to recognize and respond to emotions are as follows:

- Octavia: This robot uses the “theory of human mind” to detect the emotional state of its environment by taking into account the facial expressions, voice, and behavior of the people. These robots can also respond with the like facial manners and behavior [1].
- Nao: This is another emotion detection robot that detects emotions using image analysis techniques and is used to treat children with autism [2, 3].
- Pepper: Pepper is another emotion detection robot of Japanese design, which identifies emotions from voice, facial gestures, and body movements. It has become popular for its interpersonal interaction and connection with the human world [4].

However, a more accurate way of capturing emotions is by tapping into the physiological features of the human body. This gives a more realistic prediction of

“feeling” experienced by people. The physiological features that have been explored are the brain signals, heart signals, the working of the lungs, the heat emitted from the body, the electrodermal activity, blood pressure, galvanic skin response, and photo plethysmography. Yuan-Pin Lin et al. [5] propose a template to optimize emotion recognition from brain signals by finding EEG signals specific to emotions and to find the most efficient classifier that gives the highest accuracy to perform this task. Reza Khosrowabadi et al. [6] employ a feed-forward neural network architecture to extract emotions from brain signals. The proposed neural network ERNN is shown to have the highest accuracy of 70.83% in arousal state and 71.43% in valence state compared to other feed-forward neural networks. Panagiotis C. Petrantonakis et al. [7] attempt to fragment signals in the time frequency domain to detect emotions. Mimma Nardelli et al. [8] describe how emotional states (triggered by sounds) can be better understood using the dynamics of the autonomous nervous system. Wanhui Wen et al. [9] use random forest on a multi-subject affective database containing 477 cases (films) for emotion recognition based on physiological changes such as fingertip blood oxygen saturation, galvanic skin response, and heart rate. Rosalind W. Picard et al. [10] propose the development of a machine that can recognize human emotions from physiological signals. To achieve this, multiple algorithms are compared for feature-based recognition of emotional states.

The above works are analyzed in the next section to understand the best method in which emotions can be detected from physiological signals. The following sections are organized as follows: Sect. 3 presents a comparison study to analyze the physiological signals and techniques used to elicit emotions, and the conclusion of the study is presented in Sect. 4.

36.3 Comparison Study

36.3.1 *Comparison of Physiological Signals to Detect Emotions*

Table 36.1 compares all physiological signals used to detect emotions. The signal with the best accuracy is analyzed for further study. The inferences from the comparison study presented in Table 36.1 are as follows:

- The least parameters used to detect emotions are PPG, GSR, and electromyogram. The classifier using which the highest accuracy is observed is the Gaussian process classifier (accuracy, 85.71%). The disadvantage of this is the process of electromyography which requires needle electrodes to be placed on skin and is generally used in clinical settings.
- Accuracies higher than 80% is seen mostly in feature sets containing EEG and ECG signals as parameters.

Table 36.1 List of physiological signals used to detect emotions

Sl. no.	Ref. No.	Parameters used	Techniques used	Accuracy
1.	[5]	EEG signals	SVM classifier	82.29%
2.	[6]	Multichannel EEG data	Feed-Forward Neural Network	Arousal-70.83% Valence-71.43%
3.	[7]	EEG signals from the left frontal area and right frontal area, asymmetrical effects	Support Vector Machine classifier	60–80%
4.	[8]	HRV obtained from ECG signals	Quadratic discriminant classifier	Valence-84.72% Arousal-84.26%
5.	[9]	Fingertip blood oxygen saturation, galvanic skin response, heart rate	Random forest classifier	74%
6.	[10]	Blood volume pressure, heart rate, respiration, facial electromyogram, skin conductance	Sequential floating forward search and Fisher Projection	81%
7.	[11]	Blood volume pressure, heart rate, Galvanic skin response, respiration	Artificial neural networks, support vector machine, random forest, neuro-fuzzy system	84.3%
8.	[12]	Photoplethysmogram, ECG, electro dermal activity, temperature	K-nearest neighbor, regression tree, Bayesian networks, and SVM	82.81–88.86%
9.	[13]	Electromyogram, ECG, skin conductance, respiration	Linear discriminant analysis, emotion specific multilevel dichotomous classification	95%: subject dependent; 70%: subject independent
10.	[14]	ECG, skin temperature, electro dermal activity	SVM	78.4%
11.	[15]	Galvanic skin response, heart rate, temperature	KNN, discriminant function analysis, Marquardt back-propagation	73.9–91.7% for KNN
12.	[16]	PPG, Galvanic skin response, electromyogram	Gaussian process classifier	85.71%: mono-user 85.09%: multiuser
13.	[17]	Radio signals reflected from lungs and heart	Quadratic discriminant classifier	System trained on: each subject: 87%, independently: 72.3%
14.	[18]	PPG, GSR	Random forest classifier	72.06% arousal 71.05% valence

- Eliminating EEG and ECG parameters, the highest accuracies obtained use radio signals – accuracy: 87%. But the disadvantage is that exposure of the human body to radio signals is fatal.

- Although EEG signals use a number of electrodes, it provides the best accuracy. However, the number of electrodes can be minimized to make the end product user friendly.

36.3.2 Comparison of Methods to Extract Emotions from EEG

From the above inferences, it is observed that EEG gives the maximum accuracy, in most circumstances to elicit emotions from people. Therefore, the comparison study in Table 36.2 investigates the various methods used to extract emotions from EEG signals. The inferences obtained from the study made in Table 36.2 are as follows:

- The fewest number of electrodes used is discussed in [25] and gives an average accuracy of 87%.
- [25] shows the best accuracy among all the papers taken for the survey. However, the number of emotions elicited is only four, and the number of subjects tested is 10, which is a low number.
- Most of the papers used in the survey use Support vector machines to get the best accuracy.
- It is observed that using a support vector machine with a 10 cross validation gives better results than a 5 cross validation although the accuracy is significantly higher in the latter. The reason for this conclusion is that the former uses 5 pairs of channels and 100 pictures as opposed to the 12 pair of channels used by [21].
- It can be inferred that using a dataset gives a reduced accuracy when compared to real-time acquisition of EEG signals.
- Another observation is that comparing the accuracy of probabilistic neural network and support vector machine, it is seen that the former gives a better accuracy despite using a preprocessed dataset and a reduced number of channels.

36.4 Conclusion

From the above survey, it can be concluded that emotion detection is a complex process and many parameters must be analyzed to provide an optimized solution for a better customizable system that can make intelligent interactions with the human world.

Table 36.2 Comparison of various mechanisms to elicit emotions from brain signals

Sl. no.	Ref. No.	Subjects tested	Emotion induction	Emotions	Dataset used	Channels used	Classification algorithm	Best accuracy
1.	[19]	32	40 music videos		DEAP dataset	9 channels – valence 8 channels – arousal	Probabilistic neural network	80.02%-valence, 80.43%-arousal
2.	[20]	11	100 pictures from Geneva Affective Picture Database	Valence-arousal	Real time	5 pairs of channels (frontal) F7-F8, AF3-AF4, F3-F4, FC5-FC6, T7-T8	Gaussian SVM with tenfold cross validation	85.41%
3.	[21]	12 (6 males, 6 females)	15 movie clips	Neutral, sad happy, disgust, tensed	Real time	12 pairs of channels used: F1-FP2, FT7-FT8, FC3-FC4, F7-F8, F3-F4, T3-T4, T5-T6, P3-P4, CP3-CP4, TP7-TP8, C3-C4, O1-O2	SVM with five-fold cross validation	93.31%-PSD, 85.39%-AI
4	[22]	13 (6 males, 7 females)	45 pictures from International Affective Picture System	Pleasant, unpleasant, neutral	Real time	16 pairs of channels used: AF3-AF4, F3-F4, F5-F6, F7-F8, FC3-FC4, FC5-FC6, C3-C4, C5-C6, CP3-CP4, CP5-CP6, P3-P4, P5-P6, P7-P8, PO3-PO4, PO5-PO6, O1-O2	SVM	82.85%
5.	[23]	22	Music videos		DEAP dataset		SVM with Leave-one-out-cross-validation	64.3%-valence
6.	[24]	8 (all females)	Audio stimuli	Positive-happy, romantic Negative-disgust, sad	Real time	7 electrode systems: F3, F4, Fz, P3, P4, T3, T4	SVM	66.2%-arousal, 68.9%-dominance, 70.2%-like
7.	[25]	10		Excited, neutral, relaxed, sad	Real time	AF3-AF4, T7-T8	Learning Vector Quantization and SVM	Excited-88%, Relax-90%, Sad-84%, Average-87%

References

1. Moshkina L, Trickett S, Trafton J (2014) Social engagement in public places. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction – HRI '14
2. Shamsuddin S, Ismail L, Yussof H, Ismarrubie Zahari N, Bahari S, Hashim H, Jaffar A (2011) Humanoid robot NAO: review of control and motion exploration. In: 2011 IEEE international conference on Control System, Computing and Engineering
3. Sung Y, Cho S, Um K, Jeong Y, Fong S, Cho K (2013) Human-robot interaction learning using demonstration-based learning and Q-learning in a pervasive sensing environment. *Int J Distrib Sens Netw* 9:782043
4. Pandey A, Gelin R (2018) A mass-produced sociable humanoid robot: pepper: the first machine of its kind. *IEEE Robot Automat Mag* 25:40–48
5. Lin Y, Wang C, Jung T, Wu T, Jeng S, Duann J, Chen J (2010) EEG-based emotion recognition in music listening. *IEEE Trans Biomed Eng* 57:1798–1806
6. Khosrowabadi R, Quek C, Ang K, Wahab A (2014) ERNN: a biologically inspired feedforward neural network to discriminate emotion from EEG signal. *IEEE Trans Neural Netw Learn Syst* 25:609–620
7. Petrantonakis P, Hadjileontiadis L (2012) Adaptive emotional information retrieval from EEG signals in the time-frequency domain. *IEEE Trans Signal Process* 60:2604–2616
8. Nardelli M, Valenza G, Greco A, Lanata A, Scilingo E (2015) Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Trans Affect Comput* 6:385–394
9. Wen W, Liu G, Cheng N, Wei J, Shangguan P, Huang W (2014) Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Trans Affect Comput* 5:126–140
10. Picard R, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans Pattern Anal Mach Intell* 23:1175–1191
11. Katsis C, Katertsidis N, Fotiadis D (2011) An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomed Signal Process Control* 6:261–268
12. Rani P, Liu C, Sarkar N, Vanman E (2006) An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Anal Appl* 9:58–69
13. Kim J, Andre J (2008) Emotion recognition based on physiological changes in music listening. *IEEE Trans Pattern Anal Mach Intell* 30:2067–2083
14. Kim K, Bang S, Kim S (2004) Emotion recognition system using short-term monitoring of physiological signals. *Med Biol Eng Comput* 42:419–427
15. Lisetti C, Nasoz F (2004) Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP J Adv Signal Process* 2004
16. Fleureau J, Guillotel P, Huynh-Thu Q (2012) Physiological-based affect event detector for entertainment video applications. *IEEE Trans Affect Comput* 3:379–385
17. Zhao M, Adib F, Katabi D (2018) Emotion recognition using wireless signals. *Commun ACM* 61:91–100
18. Ayata D, Yaslan Y, Kamasak M (2018) Emotion based music recommendation system using wearable physiological sensors. *IEEE Trans Consum Electron* 64:196–203
19. Zhang J et al (2016) PNN for EEG-based emotion recognition. In: 2016 IEEE international conference on Systems, Man, and Cybernetics (SMC)
20. Jatupaiboon N, Pan-ngum S, Israsena P (2013) Emotion classification using minimal EEG channels and frequency bands. In: The 2013 10th international joint conference on Computer Science and Software Engineering (JCSSE)
21. Liu S, Meng J, Zhang D, Yang J, Zhao X, He F, Qi H, Ming D (2015) Emotion recognition based on EEG changes in movie viewing. In: 2015 7th international IEEE/EMBS conference on Neural Engineering (NER)
22. Liu S, Zhang D, Tong J, He F, Qi H, Zhang L, Ming D (2017) EEG-based emotion estimation using adaptive tracking of discriminative frequency components. In: 2017 39th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)

23. Naser D, Saha G (2013) Recognition of emotions induced by music videos using DT-CWPT. In: 2013 Indian Conference on Medical Informatics and Telemedicine (ICMIT)
24. Paul S, Mazumder A, Ghosh P, Tibarewala D, Vimalarani G (2015) EEG based emotion recognition system using MFDFA as feature extractor. In: 2015 international conference on Robotics, Automation, Control and Embedded Systems (RACE)
25. Djamal E, Lodaya P (2017) EEG based emotion monitoring using wavelet and learning vector quantization. In: 2017 4th international conference on Electrical Engineering, Computer Science and Informatics (EECSI)

Chapter 37

A Smart Agricultural Model Using IoT, Mobile, and Cloud-Based Predictive Data Analytics



P. Anand Prabu and L. S. Jayashree

Abstract In recent times, the amount of data generated by the IoT devices is very huge and the traditional databases do not have enough storage space. So the need for cloud storage becomes essential. Data mining techniques are used to analyze this big amount of data available in cloud. The study of smart agriculture system has cloud-based data analytics with IoT as a major role. The role of Information and Communication Technologies in the field of smart agriculture model is very important to extract the information from the field. In this paper, the IoT device records the data from the agriculture field and stored in the cloud database. Data analysis is done on the data available in cloud, and based on the data mining technique used, the prediction is performed. The predicted information is sent to the farmer through a mobile phone application. The main aim is to increase the production and reduce the cost of the products based on the predicted information.

Keywords IoT · Big data · Map reduce · WSN · Predictive analytics · Cloud computing · Hadoop · Mobile computing · Data mining

37.1 Introduction

In the past years, applications are being developed using the recent technologies like Internet-of-Things, Big Data, cloud, and mobile computing. Nowadays, the smart world concepts such as smart cities, smart homes, and smartphones have emerged. Agriculture is the backbone of the Indian economy. Due to the lack of maintenance, crops get damaged which causes a huge loss for farmers. So the smart agriculture

P. Anand Prabu (✉)

Department of Computer Science and Engineering, Chettinad College of Engineering and Technology, Karur, Tamil Nadu, India

L. S. Jayashree (✉)

Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_37

method was introduced. The production can be increased using the latest methods in the smart agriculture model. The information about different environmental aspects is gathered through the smart agriculture method, and it helps to monitor the system. The environmental factors that affect crop production include insects, pests, climatic conditions, etc. The following technologies can be integrated together to provide solutions for the problems faced by the farmers.

1. Internet of Things (IoT)
2. Cloud computing
3. Big Data
4. Mobile Computing

37.2 Literature Review

Currently, development in the field of agriculture is mainly based on the present system models. The information that is available from the field is used currently by decision support systems (DSSs) to decide the crop sequence. As the need for this system is increasing among the farmers, advanced smart techniques need to be used to process the data. So the need for the smart agriculture model increases to address farmers' needs.

Sanjay D. Sawaitul et al. have proposed a method of predicting the climatic conditions using back propagation algorithm. The prediction was mainly based on the past recorded data. The author collected data using the sensors in the field and used the data for prediction. The parameters of data collected include direction of wind, speed, temperature, humidity, etc. Each parameter plays a significant role in the prediction process using the back propagation algorithm.

I. Jagielska et al. have proposed an easy way to predict the crop yield using the fuzzy set theory and the probability theory. Based on the climatic conditions, crop yield prediction can be done.

Ramesh D et al. have done research work on agriculture crop yield analysis using data mining techniques. Data on sowing area, rainfall, production ratio, etc., were collected from the East Godavari district of Andhra Pradesh during 1965–2009. The multiple linear regression (MLR) algorithm and K-means prediction algorithm were used to predict the crop production yield, and high accuracy and generality were achieved.

Sindhu et al. [1] have proposed a decision tree algorithm that can be used to increase the productivity. The factors that influence the prediction process are rainfall, surface temperature, and humidity. In Naive Bayesian classification algorithm, these factors are used and prediction is done.

Wen-Yaw Chung et al. have proposed a research work using sensors for recording data from the field and using the cloud storage to store the data. A system with the combination of cloud and wireless sensor is used to monitor the land from anywhere.

Wang et al. [2] have presented their work in Agro cloud which is used to store the details of farmlands, farmers, vendors, and e-governance. All the data related to agricultural products are stored in the cloud. The BeagleBone Black sensor tool kit records data from the field, and prediction is done using data mining techniques.

Duncan Waga et al. have presented their work based on environmental factors. Cloud computing provides easy access to storage for the farmers. HDFS cloud system is used to store the data. The Hadoop Distributed File System (HDFS) is a distributed file system which easily stores and retrieves data. Its services are very flexible and can be used for collecting and aggregating the data efficiently. Hadoop packages are built on a distributed cloud environment in HDFS.

Rajesh et al. proposed a novel method by using a service-oriented architecture model for integrating the sensor information and cloud computing. The data which is available in cloud storage can be accessed by the users through the cloud service provider. The cloud technology is used to handle the data from the database and used for the prediction process. The sensor that is used to record the data is integrated within the cloud model through the Internet. The information that is collected through the sensors can be used based on the requirement of the industry. The information which is available in the cloud storage is very important for the industry.

37.3 Proposed Methodology

Our world is becoming smarter because of the IoT technologies and its applications. From the literature review, it is observed that all the technologies mentioned above are used separately in predicting the yield. Now a new smart agriculture model is proposed by integrating the IoT and Mobile and cloud technologies to improve the crop yield and reduce the cost of production.

IoT leads to the development of the many applications across all domains. The IoT devices are used to record the data from the field, and prediction is performed using the predictive data mining technique. All the data that is recorded is stored in the cloud database.

The cloud database is used in the agriculture sector. Cloud based system with the mobile technology and it as follows:

- Cloud database is used to store the crop information, prices of the fertilizers, and crop prices.
- In this agriculture model, cloud computing provides flexibility, predictability, scalability, and optimization for smart agriculture.
- The information about the current market cost is shared to farmers.
- The collection of data through the IoT devices has a major role in the smart agriculture model.

In this smart agriculture model, the farmers access the cloud services through the service provider. The service provider then allocates the storage space for the farmers to store the data, and they get access to perform the analysis. Big data

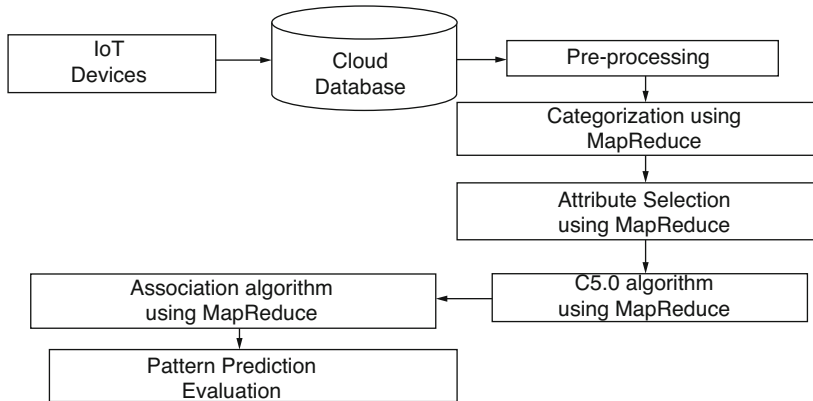


Fig. 37.1 Proposed method

technology is used in customer relationship management, fraud detection, healthcare, insurance, financial prediction analysis, and medical decision support system, and prediction by analyzing the data is done along with the help of data mining concepts.

Map Reduce is the most widely used concept to handle large amounts of data that are available in the cloud database. Also Map Reduce is used to process the data at multiple nodes simultaneously. In this model, the Map Reduce concept is used to perform the data analysis process and predicts the data based on analysis. Similarly, data mining can also be used to analyze the data, and prediction can be done based on the requirement of the end user. But data mining is not suitable for volumes of data in terabyte. For a large data set, the Map Reduce technique is the best approach for predictive analysis of the data.

Based on the literature above, the problem of analyzing the data and predicting the crop pattern was identified. The proposed solution was to collect the data from various sources and perform the data mining operations to make the data clear and available for predictive analysis. Using the Map Reduce technique, the data is separated based on the required format, and predictive analysis was done and the crop pattern can be predicted.

A smart agricultural model integrating IoT, Mobile Technology, and cloud Computing is proposed to overcome the problems identified, and the model is shown in Fig. 37.1.

37.4 Conclusion

The proposed smart agriculture model is used to decide the best crop sequence that can be followed based on the collected information using the IoT devices. Based on the information collected over the years, prediction for better production of crop and

the best crop sequence is done. Though the farmers from their past experience know the fertilizer requirements, the climatic conditions prevailing and the soil conditions also play an important role. The proposed model using the predictive data analytics technique predicts the best crop sequence, fertilizer requirements, and water requirements. All the information reaches the farmers through a mobile phone application via the Internet. Our future work will be toward interfacing different soil nutrient sensors with IoT tools, collecting and storing the data into the cloud database, and using it to analyze and predict the best algorithm that is suitable for agricultural Big Data analysis in order to get the desired outcome.

References

1. Sindhu MR, Pabshettiwar A, Ghumatkar KK, Budhehalkar PH, Jaju PV (2012) E farming. *Int J Comput Sci Inform Tech* 3(2):3479–3482
2. Wang Q, Terzis A, Szalay A (2010) A novel soil measuring wireless sensor network. *IEEE Trans Instrum Measur* 1:412–415

Chapter 38

Machine Translation Using Deep Learning: A Comparison



S. Swathi and L. S. Jayashree

Abstract In recent days, machine translation is rapidly evolving. Today one can find several machine translation systems that provide reasonable translations, although they are not perfect. The main objective of machine translation is to provide interaction among the people speaking two different languages. Machine translation, being an important task of natural language processing, leads to the development of different approaches, namely, rule-based machine translation, statistical machine translation, and neural machine translation for the translation process. The recently proposed method is the neural machine translation which improves the quality of translation between natural languages through neural networks. Neural machine translation led to remarkable improvements in the translation process by retaining the contextual information. End-to-end neural machine translation uses RNN Encoder-Decoder mechanism to train the neural translation model with bilingual corpora which is bilingual parallel sentence pairs, an important resource of machine translation. NMT has a reasonable BLEU score which is the evaluation metrics for machine translation. In this paper, we present a survey on the different kinds of machine translation approaches with their strengths and limitations and the various evaluation metrics to measure the accuracy of the translation.

Keywords Machine translation · Natural language processing · Rule-based translation · Statistical machine translation · Neural machine translation

Abbreviations

RNN	Recurrent neural network
LSTM	Long short-term memory
NMT	Neural machine translation

S. Swathi (✉) · L. S. Jayashree (✉)

Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_38

BLEU	Bilingual evaluation understudy
NIST	National Institute of Standards and Technology
TER	Translation error rate

38.1 Introduction

Machine translation is one of the most active research areas in natural language processing. Machine translation refers to the software that can perform the task of instant translation of the input from the source language to the target language in a short period of time. While performing the translation task, either human or automated, the meaning of an input sentence in the source language must be fully restored in the target language. Machine translation does not mean simply a word-for-word substitution. The translator must interpret and analyze the meaning of all the words in a sentence and know how each word influences the others. To achieve this kind of better translation, various approaches of machine translations evolved. Machine translation has been adapted which decrease the effect of language as a barrier of communication. Also the reason for choosing automatic machine translation rather than human translation is that they are better, faster, and cheaper than human translation.

Today a number of machine translation system are available that are capable of translating to a sufficient extent but not a perfect system which are used in several fields. The great challenge of machine translation is to handle various types of idioms and phrases which may vary according to the language, and it may take several trainings to improve the translation. Hence it will be very difficult for humans to design the perfect translation system. In this paper, we presented the study of various machine translation approaches that have been proposed for the task of translation.

38.2 Various Approaches for Machine Translation

38.2.1 *Rule-Based Machine Translation*

Rule-based machine translation mostly involves the dictionary creation for each word in the parallel corpus. It translates based on the syntactic and semantic rules constructed based on the linguistic knowledge of the languages used. It requires a large number of linguistic rules generated by analyzing the grammar of both the source and the target languages. This involves generating the intermediate representation from which the sentence in the target language is obtained [1]. It has the following sub-categories.

38.2.1.1 Direct Machine Translation

Direct machine translation is just a word-by-word substitution based on the bilingual-dictionary generated. This system takes the source language as an input and produces the target language as an output based on the linguistic rules in the dictionary. It does not consider the relationship of the words in the sentence and hence results in ambiguity problem.

38.2.1.2 Interlingual Machine Translation

Interlingual machine translation is not merely a dictionary mapping. This system transforms the source sentence into an interlingual language, i.e., a neural language independent of both the source and the target languages [2]. Thus the target sentences are produced from the abstract interlingual representation. But the problem is in defining such an interlingual language, and the translation is independent of target language.

38.2.1.3 Transfer-Based Machine Translation

Transfer-based system, on the analysis of source language, generates the source intermediate (bilingual dictionary) which is then transferred into the target intermediate (target dictionary), i.e., both these dictionaries are used to produce the target language [3]. This retains the meaning of the sentence to an extent, which is better than the other two rule-based machine translation systems.

38.2.2 Corpus-Based Machine Translation

Corpus-based machine translation is based on the statistical analysis of the source and the target language which requires a huge volume of data for training the system. This includes two approaches for machine translation.

38.2.2.1 Statistical Machine Translation

Statistical machine translation system performs translation better than the traditional rule-based translation system. It works based on statistics, i.e., they think in probabilities. It has different models for translation [1]. At the first stage, it generates many possible target sentences for a given input source sentence, i.e., it generates the possible translation for each word (or) phrase of a sentence and assigns the probability for each translation. The language model computes the probability of the target

language $P(T)$. In the next stage, it selects the best candidate based on the scoring function. The translation model computes the probability of the translation for the target sentence given the source sentence $P(T|S)$ which has to be maximized results in the better translation to an extent.

$$P(T|S) = \operatorname{argmax} P(T) * P(S|T) \quad (38.1)$$

The statistical machine translation system generates many possible translations of a word and ranks them rather than simple word-by-word substitution. This reduces the ambiguity of a sentence but hugely depends on the bilingual corpora.

38.2.2.2 Example-Based Machine Translation

An example-based machine translation system has two modules: retrieval module and an adaptation module. The input source for the translation process is the bilingual corpus. The retrieval module retrieves a sentence from the input bilingual corpus which is similar to the given source sentence along with its translation. The adaptation module then adapts the retrieved translation to get the final corrected translation. It depends on the similarity and hence is suitable for the similar structure languages.

38.2.3 Neural Machine Translation

The recently used machine translation approach is the neural machine translation which depends on purely deep neural networks. This system uses the RNN Encoder-Decoder LSTM model to translate the source language to the target language which will retain the contextual information of the sentence [4].

The RNN Encoder-Decoder model consists of two recurrent neural networks (RNN). Among them, one does the encoding process and the other the decoding which are required for the translation. The encoder maps a variable-length input in a source language to a fixed-length vector, and the decoder maps the vector representation back to a variable-length output in a target language.

LSTM Neural Networks, which stand for Long Short-Term Memory, are a particular type of recurrent neural networks that have some internal contextual state cells that act as long-term or short-term memory cells. Machine translation needs to depend on the historical context of inputs to retain the meaning of the given sentence, rather than only on the very last input. RNN Encoder-Decoder together with LSTM captures both semantic and syntactic structures of the phrases from the source sentence to restore the contextual information [5].

Let S and T represent the source and target words of those sentences, respectively. The encoder RNN converts the source sentences S_1, S_2, \dots, S_M into vectors of fixed dimensions. The decoder outputs one word at a time using conditional probability

$$P(T|S) = P(T|S1, S2, S3, \dots, SM) \quad (38.2)$$

that is, the occurrence of a word depends on the sequence of previous input words to restore the meaning of the given sentence [6].

38.3 Comparison Among the Different Machine Translation Techniques

From the earlier days, machine translation has been adapted in various fields with various machine translation techniques. For instance, machine translation for watcher (MT-W) focused on the readers to gain information sourced in other languages who accepted the bad translation rather than nothing. Similarly, a type of machine translation called machine translation for translator had been proposed which focused on the human translators by providing online dictionaries. Those types of machine translation adapted different machine translation techniques to perform the task of translation. On summarizing the evolution of the different types of machine translation, the recently proposed technique called neural machine translation results in better quality of translation. The comparison of the various machine translation techniques with their advantages and disadvantages is shown in Table 38.1.

Table 38.1 Machine translation techniques comparison

Machine translation	Advantages	Disadvantages
Direct machine translation	Easy to implement	Considers only lexical words
	Effective for languages with similar grammar rules	Does not consider the relationship among the words
Interlingual machine translation	It suits only when the number of target languages is more	Defining the neutral language representation (universal abstract)
Transfer-based machine translation	It depends partially on the language pair and hence less complex	Can have loss in meaning for some source text
Statistical machine translation	No need for manual development of linguistic rules	Depend on huge amount of parallel corpus
	Generic applicability to any pair of languages	Results are not accurate
Neural machine translation	Focuses on contextual information than other MT	Make use of efficient semantic structure of data and hence the complex system

38.4 Evaluation Metrics

Evaluation of performance of the translation system by humans is not always accurate. Also it takes time and is very expensive. Hence the evaluation is done via the following metrics.

38.4.1 *Bleu Score*

The Bilingual Evaluation Understudy Score, or BLEU, is a metric for evaluating a generated sentence to a reference sentence. BLEU is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another, i.e., the accuracy of the translation. A perfect match of the output sequence with the referred sentence results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0. Few translations will attain a score of 1 unless they are identical to a reference translation, but this is not possible even with the human translation. This metric is based on n-gram score.

Ex:

System Output: "the cat", "cat is", "is here"

Reference Sentence: "the cat", "cat is", "is on", "on the", "the wall"

"the cat": 1

"cat is": 1

"is here": 0

Bigrams score = $(1+1+0)/3 = 2/3 = 0.67$

38.4.2 *NIST Score*

NIST is also the metric for evaluating the quality of machine translation based on BLEU score. Rather than calculating the simple n-gram score, this metric also calculates how informative the particular n-gram is. The high preference is given to the n-grams which are more informative. This score is designed to improve the BLEU score.

Ex: The cat sat on the wall.

The bigram "cat sat" are given more weightage than for the bigram "on the".

38.4.3 Translation Error Rate

The Translation Error Rate (TER) metric calculates the number of post-editing actions required to match the translated sentence with the reference sentence. This metric can be defined as follows:

$$\text{TER} = \frac{E}{w_R} \quad (38.3)$$

where

E denotes the minimum number of edits required for an exact match with the reference text.

w_R denotes the average length of the reference text.

38.5 Conclusion

In this paper, various machine translation techniques which are adapted to provide the automated translation process have been reviewed. Also, this paper has presented about the various evaluation metrics that can be used to estimate the quality of different automated machine translation systems. Among those machine translation systems, neural machine translation provides more accuracy with the huge bilingual corpora as LSTM cells are used which can store the meaning of the text for a period of time. This helps in restoring the contextual information in the translated text and achieves greater quality in the translation process.

References

1. Saini S, Sahula V (2015) A survey of machine translation techniques and systems for Indian languages. In: IEEE international conference on Computational Intelligence & Communication Technology, pp 675–681
2. Ashraf N, Ahmad M (2015) Machine translation techniques and their comparative study. Int J Comput Appl 125(7):25–31
3. Harjinder Kaur, Vijay Laxmi (2013) A survey of machine translation approaches. Int J Sci Eng Technol Res 2(3)
4. Zakaria El Maazouzi, BadrEddine El Mohajir, Mohammed Al Achhab (2017) A technical reading in statistical and neural machine translation system. In: IEEE international conference on Information Technology, pp 157–165
5. Xiao-Xue Wang, Cong-Huizhu, Sheng Li, Tie-Jun zhao (2016) Neural machine translation research based on the semantic vector of the tri-lingual parallel corpus. In: International conference on Machine Learning and Cybernetics
6. Nakamura N, Isahara H (2017) Effect of linguistic information in neural machine translation. International conference on Advanced Informatics, Concepts, Theory and Applications

Chapter 39

Societal Impact of Framework for Energy-Efficient Clustering Algorithms in Mobile Wireless Sensor Networks



K. Juliet Catherine Angel and E. George Dharma Prakash Raj

Abstract Mobile Wireless Sensor Networks (MWSNs) are capable of sensing various types of events and change their position frequently in a specific sensing area. Architecture, energy, mobility, degree, distance, topology, localization, and data collection are the key factors in designing an energy-efficient MWSN. The applications of MWSNs can be widely divided into single-based, grid-based, and high density-based area applications. FEECA (Framework for Energy-Efficient Clustering Algorithms) is proposed to serve the various types of MWSN applications with reduced energy consumption and increased network lifetime. This paper deals with the societal impact of FEECA in various applications.

Keywords Mobility · Sensor · Energy · Framework · Topology · Data collection

Abbreviations

MWSN	Mobile wireless sensor networks
FEECA	Framework for energy-efficient clustering algorithms
BS	Base station
CH	Cluster head
CM	Cluster member
EEECA	Enhanced Energy-Efficient Clustering Algorithm
GEECA	Grid-based Energy-Efficient Clustering Algorithm
DMGEECA	Density-based Mean Grid Energy- Efficient Clustering Algorithm

K. Juliet Catherine Angel (✉)
Department of Computer Science, Holy Cross College (Autonomous), Tiruchirappalli, India

E. George Dharma Prakash Raj (✉)
School of Computer Science, Engineering and Applications, Bharathidasan University,
Tiruchirappalli, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_39

39.1 Introduction

MWSN [1–3], a network of mobile sensor nodes, is used in a variety of real-time applications. Mobile sensor nodes consist of a microcontroller, sensors, a radio transceiver, and a battery for power supply. The small size of the battery of the sensor node is a major constraint which leads to the use of energy-efficient algorithms for sensing data and transmitting them to the base station (BS). Clustering [4] is an energy-efficient technique in which the sensing area is divided into non-overlapping groups. Each group is governed by a node referred to as cluster head (CH) with all the nodes within a cluster referred to as cluster members (CM). The cluster head (CH) collects the data from all the member nodes and aggregates them and sends it to the BS.

This paper presents a framework named FEECA (Framework for Energy Efficient Clustering Algorithms in MWSN) and its societal impact in various real-time applications. This paper is organized as follows: Section 2 provides an overview of challenges in real-time MWSN applications, Section 3 focuses on the FEECA Framework, Sect. 4 presents the services offered by FEECA, Sect. 5 depicts the societal impact of FEECA, and finally Sect. 6 concludes this paper.

39.2 Challenges in Real-Time MWSN Applications

MWSN faces a lot of challenges [3, 5, 6] in real-time applications as the nodes are in mobility. Some of the challenges are listed as follows:

39.2.1 *Localization*

The mobility of the sensor nodes leads to the localization of the nodes in the sensing region. This requires time, energy, and availability of a rapid localization service which leads to the depletion of energy and reduces the network lifetime.

39.2.2 *Network Coverage*

Network Coverage is based on the quality of service that the network provides which decreases due to the undesirable sensor deployment and sensor failures. Critical applications like disaster areas, toxic regions or battlefields, natural limitations, and external harsh environments strongly affect the lifetime of sensors.

39.2.3 *Dynamic Network Topology*

The mobility of the nodes leads to dynamic topology. In dynamic topologies, the position of the sensor nodes change with time, and route discovery must be performed repeatedly which consumes power, time, and bandwidth.

39.2.4 *Energy Consumption*

The sensor nodes are resource constrained with limited energy. Energy-efficient algorithms need to be designed to conserve energy and increase network lifetime with minimum energy dissipation and optimized communication.

39.2.5 *Mobility of Sensor Nodes and Sink*

In MWSNs, the data gathering issues with mobile BSs, mobile relay node, and mobile sensor nodes are a great challenge. The sensor control methods for delivering data to the sink efficiently considers node failures, fault tolerance and delay in data delivery to improve the efficiency of the network.

39.3 FEECA: Framework for Energy-Efficient Clustering Algorithms in MWSN

Framework for Energy-Efficient Clustering Algorithms (FEECA) is developed to integrate the proposed Energy-Efficient Clustering Algorithms to enhance the performance of MWSNs by providing reduced energy consumption and increased network lifetime. The proposed FEECA (Framework for Energy Efficient Clustering Algorithms) is given in Fig. 39.1. The topmost part of the framework is the MWSN applications. The nodes are deployed randomly in a MWSN through which the data are sensed and sent to the BS. FEECA is comprised of three types of application areas based on the real-time application and the user's requirements.

The MWSN applications can be single area-based, grid area-based, and high density-based. EEECA (Enhanced Energy-Efficient Clustering Algorithm) [7] is designed to provide an energy-efficient way to sense the data and transmit it to the BS through a CH and to increase the network lifetime during single area-based application. GEECA (Grid-based Energy-Efficient Clustering Algorithm) [8] is developed to offer an energy-efficient way to sense the data from the grid-structured area and transmit it to the BS through the CHs in each grid and to increase the

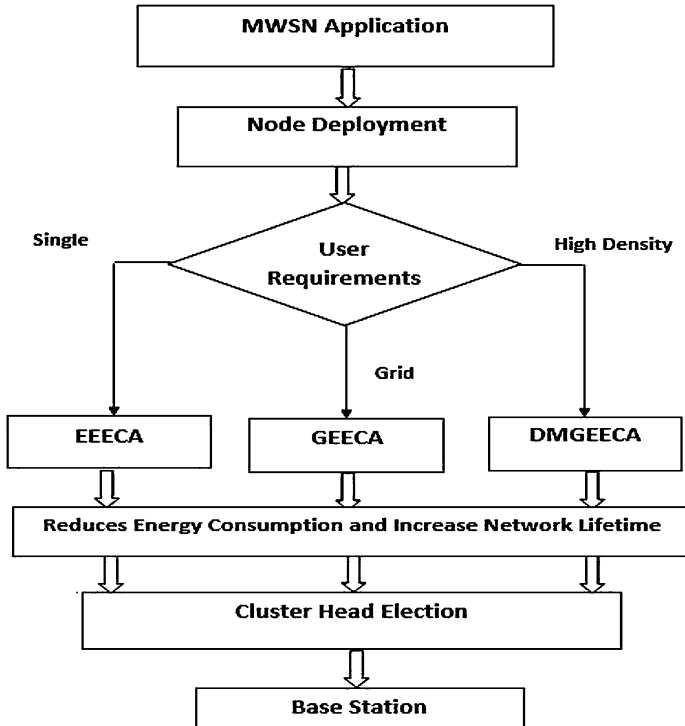


Fig. 39.1 FEECA: Framework for Energy-Efficient Clustering Algorithms for MWSN

network lifetime during grid area-based application. DMGEECA (Density-based Mean Grid Energy-Efficient Clustering Algorithm) [9] is designed to provide an energy-efficient way to sense the data from a high-density sensing area and transmit it to the BS through the CHs in each grid and to increase the network lifetime during high density-based application.

39.4 Services Offered by FEECA

The services provided by FEECA are based on single sensing area, grid-based area, and density-based area depending upon the MWSN application as shown in Fig. 39.2. Services offered by FEECA based on the proposed algorithms are:

- Single area-based services
- Grid area-based services
- High density area-based services

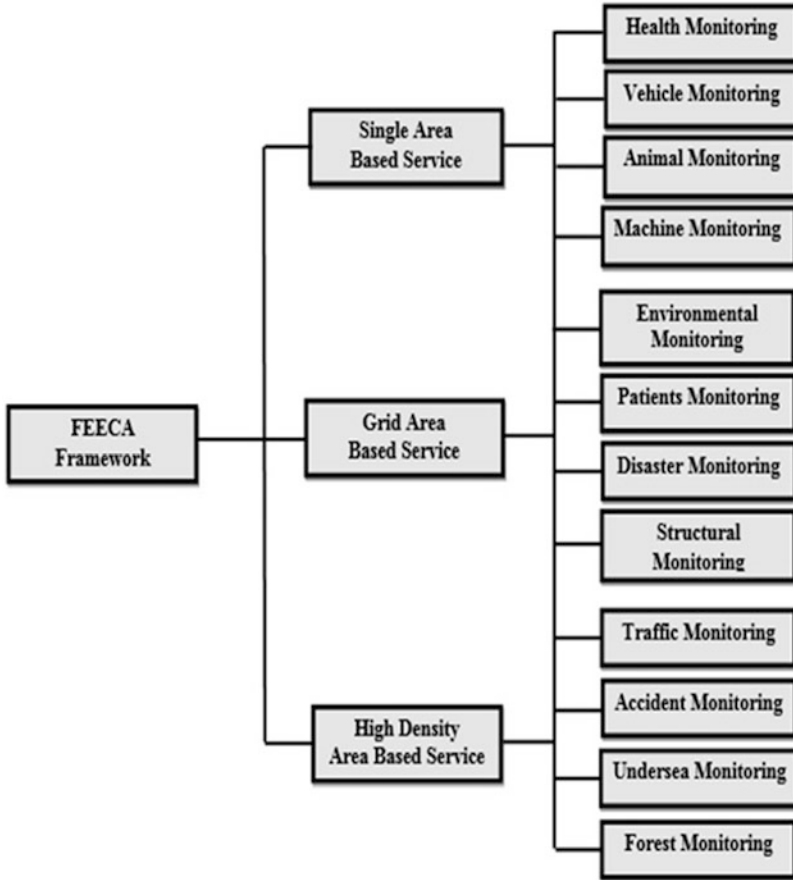


Fig. 39.2 Services offered by the FEECA Framework

39.4.1 Single Area-Based Services

This service provides data communication through clustering with the appropriate selection of the CH based on the application. The sensing and data transmission is designed for single area-based applications in order to reduce the energy consumption and increase the network lifetime.

39.4.2 Grid Area-Based Services

Grid clustering is used to sense data from a collection of areas and transmit it to the BS with reduced energy consumption and increased network lifetime. This type of services is used to monitor a very big sensing area where the area is uniformly

partitioned into grids. The sensed data from various grids are sent to the BS using grid clustering. Multiple users are monitored by integrating the collected data and forwarding it to BS.

39.4.3 High Density Area-Based Services

These types of services are designed to monitor highly densely populated area. Multiple users are monitored by using density-based clustering. The data communication of the sensed data is carried out easily by increasing the number of clusters to reduce the energy consumption and increase the network lifetime.

39.5 Societal Impact of FEECA

The various applications served by the FEECA Framework for MWSN based on the area of applications are shown in Fig. 39.2.

39.5.1 Single Area-Based Applications Served by FEECA

FEECA provides an energy-efficient clustering algorithm (EEECA) for single area-based applications. EEECA is based on energy consumption, mobility, distance, and degree of the node in order to reduce the energy consumption and increase the network lifetime for monitoring, data collection, and data transmission. The single area-based applications served by FEECA through EEECA are health monitoring, vehicle monitoring, animal monitoring, and machine monitoring.

39.5.1.1 Health Monitoring

In health monitoring, the mobile sensors are fixed on individuals to monitor the heart rate, blood pressure, blood sugar, and growth of fetus in the mother's womb, which are constantly checked and the reports are sent to the doctors. Any abnormalities are automatically intimated to the doctor and ambulance team.

39.5.1.2 Vehicle Monitoring

FEECA provides an energy-efficient clustering through EEECA for vehicle monitoring in factors like speed, acceleration, efficiency of the break, mileage, etc.

The efficiency of the parts of a vehicle is monitored using mobile sensors to decide whether the parts need to be remodeled.

39.5.1.3 Animal Monitoring

FEECA provides an energy-efficient clustering through the EEECA for animal monitoring by fixing sensor nodes on animals individually to monitor movements, behavior patterns, migration, and feeding habits, treat their health, improve breeding, and take measures to reduce extinction of wild animals.

39.5.1.4 Machine Monitoring

FEECA provides an energy-efficient clustering through EEECA for machine monitoring of the function and production capacity of various machineries that can be constantly monitored. It helps to check the capacity of each machinery and helps in replacing them as the efficiency reduces which increases the production of the industry and increases the profit.

39.5.2 Grid Area-Based Applications Served by FEECA

FEECA provides an energy-efficient clustering algorithm (GEECA) for grid area-based applications. GEECA provides an energy-efficient clustering based on grid concepts, energy consumption, mobility, distance, and degree of the node in order to reduce the energy consumption and increase the network lifetime for monitoring, data collection, and transmission. The grid area-based applications served by FEECA through GEECA are environmental monitoring, patient monitoring, disaster monitoring, and structural monitoring.

39.5.2.1 Environmental Monitoring

FEECA provides an energy-efficient clustering through GEECA for environmental monitoring for finding the climatic changes. The mobile sensors are used to monitor the light, heat, humidity, rainfall, temperature, pollution, etc., to find whether the prevailing atmosphere is fit for the normal life of the creatures.

39.5.2.2 Patient Monitoring

FEECA provides an energy-efficient clustering through GEECA for patient monitoring in a hospital where sensors automatically monitor the health conditions of the

patients and send results to the doctors for further treatments which helps in improving the health conditions of the patients even faster.

39.5.2.3 Disaster Monitoring

FEECA provides an energy-efficient clustering through GEECA for disaster monitoring. Mobile sensors are used to monitor the conditions that cause natural disasters and in finding the cause of tsunamis, earthquakes, floods, and volcanic eruptions, which helps in taking precautionary measures by giving alerts to the people.

39.5.2.4 Structural Monitoring

FEECA provides an energy-efficient clustering through GEECA for structural monitoring by fixing mobile sensors to physical structures in order to monitor the performance of structures during changes in load, temperature, rotation, displacement, inclination, strain, fatigue, vibration, deflection, and crack movement. This helps in detecting changes in the structure like cracks, vibration, and damage and replacing or repairing the flaws of the structure immediately.

39.5.3 High Density Area-Based Applications Served by FEECA

FEECA provides an energy-efficient clustering algorithm (DMGEECA) for high density area-based applications. DMGEECA provides an energy-efficient clustering based on high density, energy consumption, mobility, distance, and degree of the node in order to reduce the energy consumption and increase the network lifetime. The high density area-based applications served by FEECA through DMGEECA are traffic monitoring, accident monitoring, undersea monitoring, and forest monitoring.

39.5.3.1 Traffic Monitoring

FEECA provides an energy-efficient clustering through DMGEECA for high-density traffic area monitoring. The heavy traffic in any area can be constantly monitored through mobile sensor nodes to prevent traffic jam which helps in diverting the traffic in case of heavy traffic prediction. This automatic monitoring of the traffic prevents accident and traffic jam in the area.

39.5.3.2 Accident Monitoring

FEECA provides an energy-efficient clustering through DMGEECA for monitoring accidents in a high-density traffic area. Constant surveillance using mobile sensors helps in detecting the occurrence of any accident and intimating it to an ambulance team to provide the necessary first aid and further treatments which helps in saving lives of many people.

39.5.3.3 Undersea Monitoring

FEECA provides an energy-efficient clustering through DMGEECA for undersea monitoring in a high-density area. The undersea surveillance using mobile sensors helps in oceanographic data collection which are needed to explore, protect, and commercial exploitation of the aquatic environment which helps in exploring volcano eruption, earthquake, or tsunami in the ocean and taking precautionary measures in saving the life of aquatic creatures.

39.5.3.4 Forest Monitoring

FEECA provides energy-efficient clustering through DMGEECA for monitoring the high-density forest area using mobile sensors which helps in detecting the climatic changes to explore, protect, and commercial exploitation of the forest. It helps in exploring the changes in the climatic conditions of forest, disaster, exploitation of forest, and pollution which helps in taking precautionary measures in saving the life of living creatures in the forest.

39.6 Conclusion and Future Work

The proposed framework FEECA contains three clustering algorithms which reduces the energy consumption of MWSN and increases the network lifetime. The existing EECA-M2 algorithm is enhanced as EEECA, which provides clustering with minimized energy consumption in single-area application. EEECA is enhanced with grid concepts such as GEECA which provides less energy consumption which leads to increase in network lifetime thereby reducing the overhead of the CH (GEECA) in grid-based applications. GEECA is enhanced as DMGEECA which provides less energy consumption with density-based clustering to reduce the energy consumption during high density of nodes. The time complexity of the proposed algorithms EEECA, GEECA, and DMGEECA is found as $O(n \log m)$ based on n number of mobile sensor nodes and m number of CHs. The energy efficiency is provided by using grid clustering and density-based clustering. In the future, this research could be extended with the implementation of the proposed FEECA using clustering concepts and incorporating the cloud concepts for better efficiency.

References

1. Kumari P, Singh MP, Kumar P (2013) Cluster head selection in mobile wireless sensor networks: a survey. In: Proceedings of the international conference on Advances in Computing and Communication – ICACC 2013
2. Ramasamy V (2017) Mobile wireless sensor networks: an overview. InTech Open Science
3. Bharti D, Behniwal M, Singh AK (2013) Performance analysis and issues of mobile sensor network. *Int J Adv Res Eng Appl Sci* 2(4):58–67. ISSN:2278-6252
4. Juliet Catherine Angel K, George Dharma Prakash Raj E (2017) Clustering algorithms in mobile wireless sensor networks – a survey. *Int J Eng Res Appl* 7(12 (Part -6)):17–21. ISSN:2248-9622
5. Chunsheng Zhu, Lei Shu, Takahiro Hara, Lei Wang, Shojiro Nishio (2010) Research issues on mobile sensor networks. ResearchGate
6. Manisha, Sharma SP (2015) Energy efficiency communication protocol with mobility in wireless sensor network: a survey. *Int J Comput Sci Inform Technol* 6(1):737–740. ISSN:0975-9646
7. Juliet Catherine Angel K, George Dharma Prakash Raj E (2017) EEECA: enhanced energy efficient clustering algorithm for mobile wireless sensor networks. In: International conference on World Congress on Computing and Communication Technologies (WCCCT), pp 267–270. <https://doi.org/10.1109/WCCCT.2016.71>
8. Juliet Catherine Angel K, George Dharma Prakash Raj E (2017) GEECA: grid based energy efficient clustering algorithm for mobile wireless sensor networks. *Saudi J Eng Technol (SJEAT)* 2(12):457–464. <https://doi.org/10.21276/sjeat.2017.2.12.1>. ISSN:2415-6264 (Online)
9. Juliet Catherine Angel K, George Dharma Prakash Raj E (2018) DMGEECA: density based mean grid energy efficient clustering algorithm for mobile wireless sensor networks. *Int J Comput Sci Eng* 6(3):27–33. E-ISSN:2347-2693. (UGC Approved Journal)

Chapter 40

Energy Demand Prediction Using Linear Regression



T. Manojpraphakar and Soundarrajan A

Abstract Big Data analytics is the latest emerging technology that requires deep knowledge in business intelligence, machine learning, and statistical methods and in deep learning. It focuses on the application of data analytics for energy demand management using real-time data. The data is then analyzed for clustering, demand forecasting, pricing, and energy generation optimization. It represents a method to predict energy usage, based on real-time data obtained from TANGEDCO-CBE, using the linear regression model (LR). The final linear regression models developed were based on daily sustained demand and consumption by comparing actual and predicted energy usage models can predict with acceptable errors. Normally the energy requirement and industrial demands are high; hence the application of these energy Big Data analyses significantly improves efficiency and provides new business opportunities.

Keywords Big data · Visualization · Energy prediction · Linear regression · Simple linear regression

Abbreviations

LR	Linear regression
HDFS	Hadoop distributed file system
IT	Information technology
SVM	Support vector machine
TANGEDGO	Tamil Nadu Generation and Distribution Corporation

T. Manojpraphakar (✉)
Faculty of Information and Communication Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India

Soundarrajan A
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India

AutoDR	Automated demand response
MBCx	Monitoring-based commissioning
GPRS	General packet radio service

40.1 Introduction

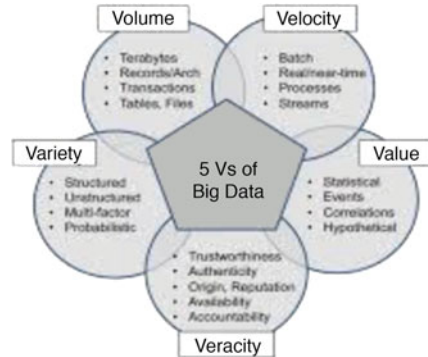
Big Data is a collection of large volumes of data that contains structured, unstructured, and semi-structured data, which is difficult to store, analyze, share, visualize, and manage with most traditional database and software techniques. Advancement in development, increasing need on platforms and in many software industry applications to handle the accuracy, rate at which enterprises remain to face in a competitive global market world.

Hadoop is an open-source software which is developed using java programming that helps in accessing the large datasets in a distributed computing model. Both the Hadoop distributed file system(HDFS) and Mapreduce are designed to work in system failures. It is developed and managed by Apache Hadoop. The Hadoop framework uses the Mapreduce algorithm that helps the data to be analyzed in parallel. It stores any type of data in its format, performs the analysis, and changes the data. Hadoop stores the information ranging from terabyte to even petabytes of data. Communication technologies have enabled the stakeholders in the energy industry to collect large amounts of useful and highly granular energy usage data. This data is generated in large volumes, and depending on its purpose, it is generated in a variety of different formats which is collected with the help of systems. The volume and diversity of data also increase with time. All these data characteristics refer to the application of Big Data.

Big Data can be characterized in 5 Vs:

- *Volume*: It represents the size of Big Data and whether the data is considered to be Big Data or not.
- *Velocity*: It refers to the speed at which the data is accessed, stored, and analyzed.
- *Variety*: Variety of data consists of structured, unstructured, and semi-structured data.
- *Value*: It represents cost of implementing the information technology infrastructure that turns the output as business value.
- *Veracity*: Since the volume of data is high, we might not assure that all the data will be accurate. So the accuracy of the data depends on it (Fig. 40.1).

The analysis was performed on two datasets. The first dataset contained daily sustained demand in MW for the Coimbatore region. The second dataset contained daily consumption in MW for the Coimbatore region. The data was analyzed to discover daily usage trends for different energy prediction models to forecast future consumption on the basis of previous usage. One such model used in this paper is linear regression model.

Fig. 40.1 5Vs of Big Data

The analysis described in this paper uses real-time data obtained from TANGEDCO-CBE and data visualization software SAP LUMIRA to transform the massive quantity of data, cluster the time series, apply time series analysis, and then draw both predictive analytics and business conclusions from the data.

40.2 Literature Survey

Kaile Zhou et al. [1] have investigated that large amounts of data are rapidly increasing day by day in the energy sector, namely, electricity data, consumption data, user interaction data, etc., in order to manage the data efficiently. Big Data technologies is used to predict the future data by data collection, data integration, and data preprocessing techniques. It is made up of three aspects, namely, power generation, renewable energy, and demand side energy management. When compared with other sectors, the energy area is facing problems in reliability, efficiency, and cost control. Based on the collection of energy data and the use of advance analytics techniques, the cost of energy consumed in various industries can be reduced. The Information Technology (IT) infrastructure needs to be improved in the area of network, storage capability, visualization, and a data processing area of energy management.

Panagiotis et al. [2] have enhanced that a smart grid is used to manage electricity demand in a sustainable manner by applying digital information and communication techniques, design and development of algorithms, and methods from large-scale datasets. Data mining is used to process the useful information from the stream of data, namely, user's consumption data and electricity data, into further usage. Dimensionality reduction is used to eliminate the unwanted attributes that are generated by smart meters.

Kun Wang et al. [3] proposed an efficient algorithm to predict the electricity price forecasting. Normally electricity price is influenced by fuel price, electricity requirement, renewable energy supply, etc., since the electricity price changes frequently and a huge amount of meters monitor fuel power generation, wind generation, and transmission in real time. An electricity price forecasting consists of two stages of

feature processing and advanced Support Vector Machine(SVM) classifier. Date-wise prediction is analyzed and proposed and implemented on parallelized and distributed systems.

Jungin Choi et al. [4] developed a model platform for a demand-side energy management. The main goal of the system is to save energy and analyze them using Big Data. The main applications are automated demand response (AutoDR) and monitoring-based commissioning (MBCx). It uses two types of Big Data clusters, namely, Hadoop cluster and spark cluster. Software modules help to collect and generate the energy data with the use of application programming interfaces.

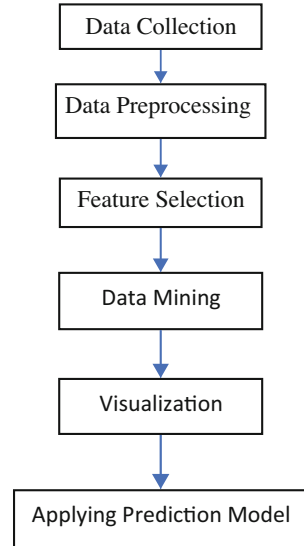
Shyam R et al. [5] studied that data collected from different sensors of a smart grid is huge in a very limited period of time, which can be analyzed and predicted using Big Data processing and various machine learning algorithms. The Big Data analytic platforms are designed with huge power and flexibility to meet all such user requirements. Apache Spark is efficiently used for the applications like real-time price forecasting, automatic demand response systems, peak time load balancing, fault identification, and online grid monitoring.

40.3 Sources of Energy Big Data

Entities need more energy than ever before, and they want it at prices more affordable than ever. Big Data allows companies to gather, store, and analyze extremely large (terabytes and petabytes) amounts of information. There are two ways that companies can implement Big Data. First, they can choose to install the architecture, or they can use Big Data stored in the cloud and tell the cloud provider to take care of the infrastructure. The data in the energy sector is collected through smart energy meters. The substation data concentrator unit will acquire data from feeder meters and will transmit the same to central data center server(s) through GPRS/EDGE. The meter data from all distribution transformers as well as HT/select LT consumers will be transmitted to the central data center server(s) through GPRS/EDGE. A PC will be installed at each substation with a purpose for local viewing and monitoring of feeders by substation staff. The substation meters are fitted with RS 485 ports, whereas meters at DT and select consumer locations fitted with RS 232 ports. As most of the meters provided presently are Modbus compliant, Modbus/TCP shall be the preferred mode of communication. The proposed Data Logging system is expected to provide continuous online monitoring and logging of essentially, though not limited to, the following parameters in respect of distribution transformers and consumers on real-time basis:

- Voltage, phase to phase and phase to neutral
- Current on each phase
- Power factor
- Frequency
- Power – Active/reactive/apparent
- Energy – Active, reactive, and apparent (Fig. 40.2)

Fig. 40.2 Process model of Big Data-driven smart energy management



40.4 Visualization Process

SAP Lumira 2.0 is a software which can have access to data sources like Excel, relational data like Oracle, IBM DB2, SAP HANA, or Big Data like Amazon Redshift, Apache Spark, or Cloudera Impala. It can manipulate, cleanse, and join data from these datasets. It creates visualizations, stories, and infographics. Finally, it can share the resulting LUMX file with other users of SAP Lumira 2.0 discovery via email. Another Add-ins XLMINER is also used. This Add-in can be used with MS-EXCEL software.

The two types of analysis are univariate analysis and bivariate analysis which is also discussed in this paper. Univariate analysis explores variables (attributes) one by one. Variables could be either categorical or numerical. There are different statistical and visualization techniques of investigation for each type of variable. Bivariate analysis is the simultaneous analysis of two variables (attributes). It explores the concept of relationship between two variables, whether there exist any association and the strength of this association or whether there are differences between two variables and the significance of these differences.

In Figs. 40.3 and 40.4, a histogram of sustained demand and consumption is shown. From the plots shown in both figures are various analysis results such as mode, variance, etc. Var. 1 in Fig. 40.3 represents the sustained demand and var. 1 in Fig. 40.4 represents the consumption. This visualization was done with the help of MS-EXCEL add-ins XLMINER.

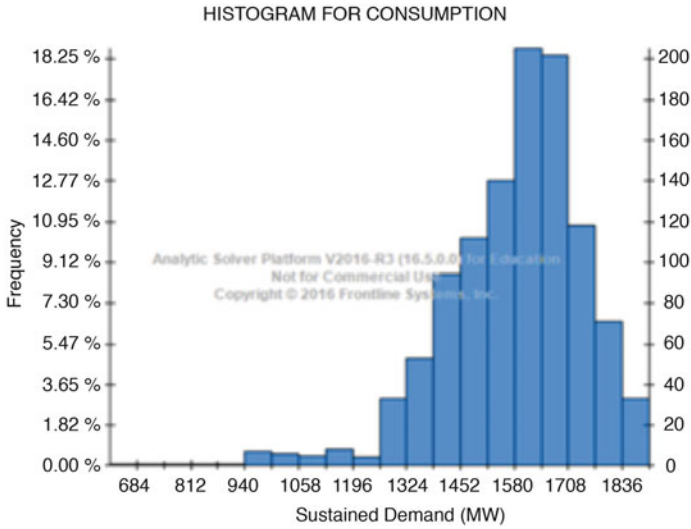


Fig. 40.3 Histogram for Sustained Demand

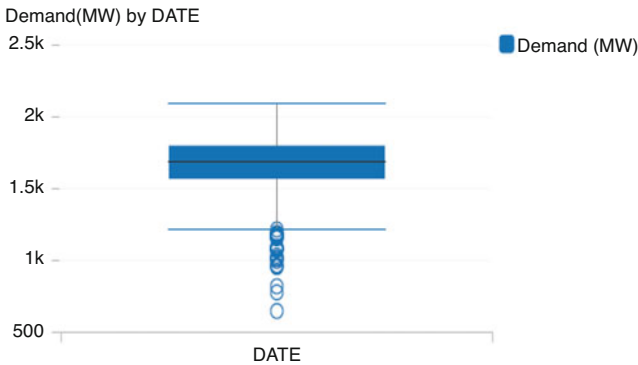


Fig. 40.4 Box plot for Sustained demand

From the above plot, 18.25% of the values, i.e., more than 200 data values, are between 1550 and 1695 MW. The distribution of data between various ranges can be found from the above histogram plot. Similarly, about 35% of the consumption lies between 1580 and 1700 MW, one-third of the data.

In Figs. 40.4 and 40.5, box plots for sustained demand and consumption are shown. From these two plots, we can find the mean, maximum, and minimum range, etc. The box plots were obtained using SAP Lumira 2.0.

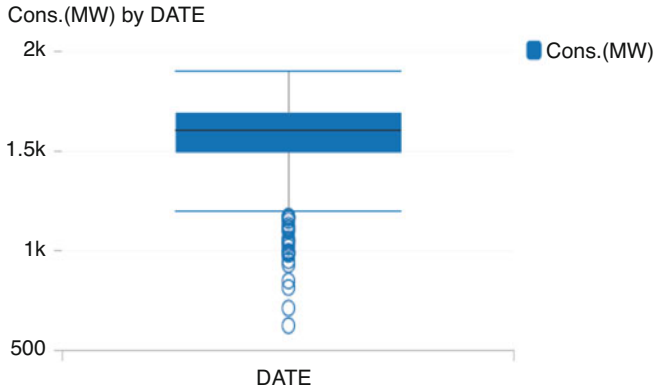


Fig. 40.5 Box plot for consumption

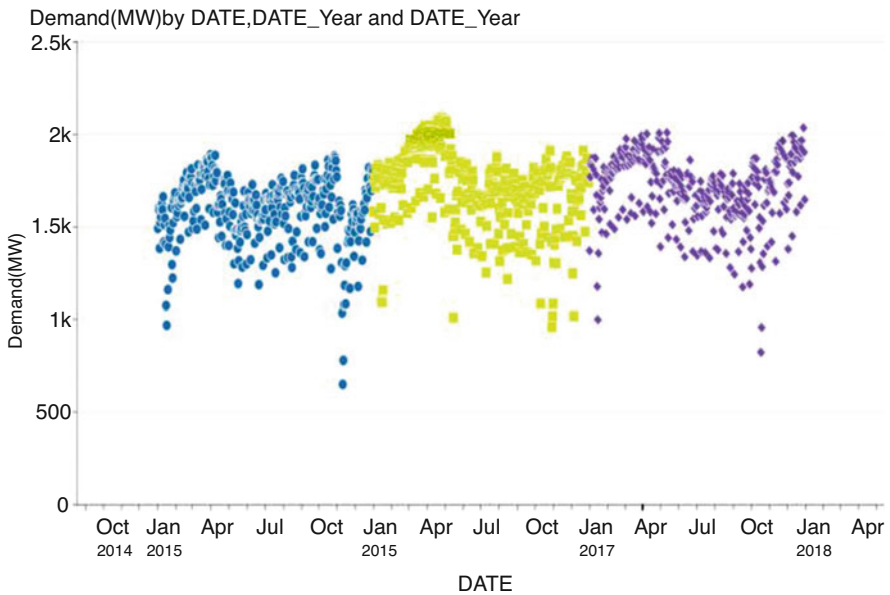


Fig. 40.6 Plot showing sustained demand and date

From this figure, the maximum demand is 2091 MW and the minimum demand is 1222.23. The mean value of the data is 1668.36 MW.

The maximum and minimum demands for the above figure are 1900 MW and 1197.6 MW. The mean for the data is 1581.352 MW (Fig. 40.6).

Next, with the help of a scatter plot, bivariate analysis was performed, but before doing so, we need to determine the linear correlation between sustained demand and

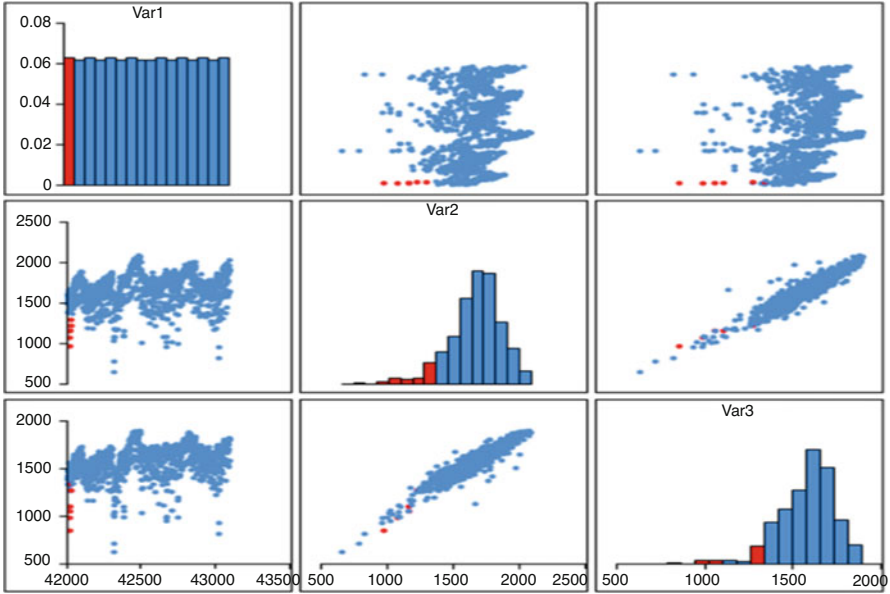
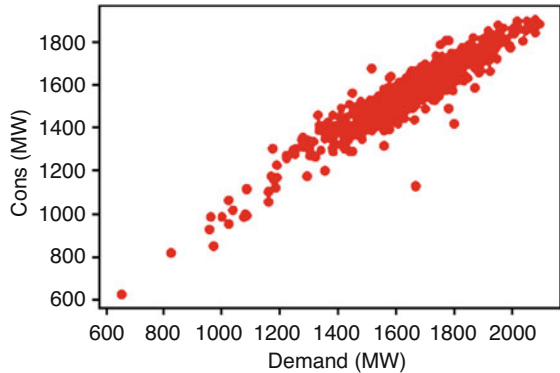


Fig. 40.7 Scatter plot matrix

Fig. 40.8 Scatter plot



consumption. The value lies between -1 and 1 , and in our case the value is 0.94 . We got the scatter plot using the SAP Lumira 2.0 software, and by using MS-EXCEL add-ins, XLMINER scatter plot matrix was formed. Figure 40.7 represents the scatter plot. The scatter plot matrix is shown in Fig. 40.8; var 1, 2, and 3 represent date, sustained demand, and consumption. Data for the highlighted region is also shown in Fig. 40.8.

40.5 Linear Regression Predictive Modeling

The chosen and developed model for forecasting energy consumption is simple linear regression. Fitting a straight line is a best fit which minimizes least squared error in approximation. It has been widely used in energy modeling projects. When compared with nonlinear models, linear regression models are practical and less tedious for solving the problems. The estimation of squared error is given by the equation

$$J = \sum [Y_i - f(x_i)]^2 \quad (40.1)$$

The equation that represents the relationship between variables Y and X is

$$Y = mx + b \quad (40.2)$$

In the above equation, Y is the dependent variable and the predicted variable X is the independent variable, i.e., the variable we are using to make predictions. The slope of the regression line is represented by m . It shows the effect that one variable has on the other. In a SLR, we derive a model based on the slope and Y -intercept derived from the data; furthermore, it is not a requisite that the variables should be exactly linear. SLR models also include the errors in the data also known as residuals; it is the difference between the practical value of Y and the predicted value. In linear regression, it is essential to note that we are trying to predict a continuous variable. In a regression model, by finding the “line of best fit,” we are able to reduce the errors.

For modeling linear regression, python is used that provides many built-in library functions that are installed

by using the pip command

Matplotlib: It is a python library used to plot 2D graphs and charts. We will be using it for representing our model.

NumPy: It is a package which is used to carry out scientific computation in python.

We will be using it for making arrays and matrix for our dataset and carry out basic computation on them.

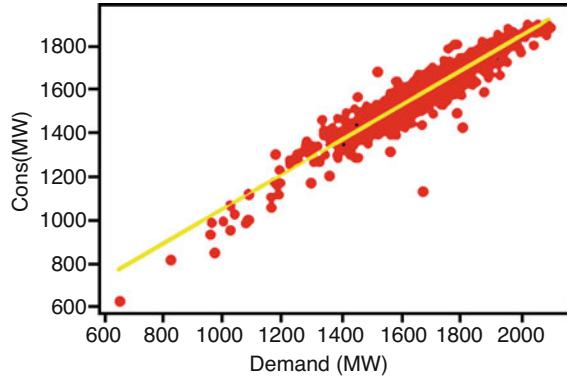
Sklearn linear_model: We will be using the linear model of sklearn to carry out our linear regression on the given dataset.

Data are divided into two types to train and test for building the model:

Training Dataset: The sample of data used to fit the model and parameters to estimate the relationship between the training data.

Test Dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset. It is independent to training dataset and includes some probability functions in the training dataset.

Fig. 40.9 Simple linear regression model



In this model, 70% of the values are randomly selected as train data and the remaining as test data.

40.6 Results

The development of the models was simple and accurate compared to actual and predicted energy usage. The final linear models were very simple (two variable regression models) compared with complex nonlinear modeling methods and can even be used in MS Excel (Fig. 40.9).

40.7 Conclusion and Future Works

This paper presents a simple linear regression (SLR) model for predicting the energy consumption for future. Using this model, we can predict values with an accuracy rate of $\pm 1.2\%$. To improve the level of prediction, some additional datasets such as voltage, power factor, temperature, etc., are used. And the same dataset implementation can be carried out using artificial neural network and time series analysis.

References

1. Zhou K, Fu C, Yang S (2016) Big data driven smart energy management: from big data to big insights. *Renew Sustain Energy Rev* 56:215–225
2. Diamantoulakis PD, Kapinas VM, Karagiannidis GK (2015) Big data analytics for dynamic energy management in smart grids. *Big Data Res* 2(3):94–101
3. Wang K, Xu C et al (2017) Robust big data analytics for electricity price forecasting in the smart grid. *IEEE Trans Big Data* 5(1):34–45

4. Jungin Choi, Mijeom Kim, Jaeweon Yoon (2015) Implementation of the big data management system for demand side energy management. In: 2015 IEEE international conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp 1515-1520. <https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.227>
5. Shyam R, Bharathi Ganesh HB et al (2015) Apache spark a big data analytics platform for smart grid. Smart Grid Technol 21:171–178

Chapter 41

Risk Prediction Analysis of Cardiovascular Disease Using Supervised Machine Learning Techniques



A. Ishwarya and S. K. Jayanthi

Abstract The best thing to avoid strategic human death rates due to curable diseases is to detect them early and prevent their onset. Presently, in our society, large numbers of death rates are due to cardiovascular disease (CVD). Hence early detection of CVD is critical even though many more practices exist for earlier prediction of risk. One approach for early disease risk prediction is the use of risk prediction models developed using machine learning techniques. These models will provide clinicians to treat heart disease of the patient in a better way. Consequently in this chapter, classification mechanisms have been applied to predict the status of the disease. The machine learning algorithms involved in the prediction of CVD are EDC-AIRS, Decision Tree, and SVM. The heart disease dataset from UCI repository has been used in this study. The predictions are denoted by means of accuracy, whereas the performance measures have been calculated in terms of sensitivity, specificity, and F-measure. Results indicate that the prediction model developed using the SVM algorithm is capable of achieving high sensitivity, specificity, balanced accuracy, and F-measure. Further, these models can be integrated into a computer-aided screening tool which clinicians can use to predict the risk status of CVD after performing the necessary clinical assessments.

Keywords Cardiovascular disease · Classification · Clinical risk prediction

Abbreviations

CVD	Cardiovascular disease
EDC-AIRS	Evolutionary data-conscious artificial immune recognition system
KNN	K-Nearest Neighbor
SVM	Support vector machines
DT	Decision tree

A. Ishwarya (✉) · S. K. Jayanthi
Department of Computer Science, Vellalar College for Women, Erode, Tamil Nadu, India

HDL	High-density lipoprotein
LDL	Low-density lipoprotein
CP	Chest pain
FBS	Fasting blood sugar
IG	Information gain
GI	Gini Index
CHOL	Cholesterol level
TRESTBPS	Resting blood pressure

41.1 Introduction

Cardiovascular disease (CVD) is one of the leading causes of death in many developed countries. CVD generally refers to the condition of narrowed or blocked blood vessels that can lead to heart attack, chest pain, or stroke. Other heart conditions, those that affect heart muscle and valves [1], are coronary artery disease, high blood pressure, cardiac arrest, peripheral artery disease, congestive heart failure, arrhythmia, and congenital heart disease. The traditional risk factors or biomarkers for predicting CVD outcomes include age, sex, systolic blood pressure, smoking habits, diabetes, total cholesterol, high-density lipoprotein (HDL) cholesterol, low-density lipoprotein (LDL) cholesterol, diabetes mellitus, family history, etc. There are other severe cardiovascular outcomes, such as HF, a kind of common disease, which can cause leg swelling, shortness of breath, and other severe symptoms [1]. Heart failure (HF) can be caused by other diseases, such as rheumatic heart disease, anemic heart disease, and toxic heart disease. Traditional risk factors can explain only half of the incidence of CVD. Therefore further efforts are needed to improve the performance of risk prediction models [2].

In recent times, a large number of research works are conducted for improving the prediction performance of CVD. Ultrasound is a widely used imaging modality, as it captures moving images, and the image features correlate well with results obtained from other imaging methods. But it takes time to read ultrasound images because the relationship between the image and tissue composition is quite complex [3]. Genetic contributions to CVD also result poor. Therefore data mining techniques can be used for earlier risk prediction. The advantages of using data mining techniques involve the capability of dealing with information, solving non-trivial problems, and producing data-driven prediction models. Some data mining techniques generate models of the data in terms of production rules, and cardiologists may then analyze them.

The risk prediction tools can give clinicians the ability to anticipate and treat CVD early, which can be more beneficial for their patients [4]. To this end, we investigate the possibility of customizing CVD risk prediction models to better meet the patient's issues and clinician's expectation. In this chapter, the well-known machine learning algorithms such as evolutionary data-conscious artificial immune recognition system (EDC-AIRS), decision tree, and SVM algorithms were explored.

The input dataset has been collected from UCI repository [1]. The patients aged 29–77 were studied [5]. Further, wide range of clinical measurements and risk factors are observed by dividing the risk classes broadly. It gives a valuable source of information for this chapter. The rest of the chapter is organized as follows. Section 41.2 discusses the literature survey, whereas Sect. 41.3 provides the details of the proposed system which includes the machine learning algorithms involved in risk prediction. In Sect. 41.4, experimental results are analyzed. Section 41.5 concludes the work.

41.2 Literature Survey

Many papers were reviewed for the improvement of the research. Some of them are explained under this literature survey section. Darwin Tay et al. [4] have developed risk prediction models for a particular heart disease, namely, myocardial infarction, based on the patient's sample age and prediction resolution. This paper deals with EDC-AIRS and SVM algorithms. They concluded that SVM is more accurate than EDC-AIRS in predicting the risk status of myocardial infarction.

Olga Troyanskaya et al. [6] deal with a comparative study of several methods for the estimation of missing values in gene microarray data. They implemented and evaluated three methods, namely, Singular Value Decomposition (SVDimpute), weighted K-Nearest Neighbors (KNNimpute), and row average. They concluded that KNNimpute is a more healthy and sensitive method for estimating missing values than SVDimpute and row average.

Darwin Tay et al. [7] proposed a new algorithm called evolutionary data-conscious artificial immune recognition system. This algorithm is implemented on four widely used benchmark datasets and established promising results. It signifies the importance of integrating the immune mechanism as a part of the learning process.

Edgar Osuna et al. [8] deal with training support vector machines algorithm for large datasets. They presented the experimental results by implementing SVM on a face detection problem dataset that consists of 50,000 data.

Jose M. Jerez et al. [9] deal with the missing data imputation task. In this work, imputation methods based on statistical techniques and machine learning algorithms were applied to a particular dataset. Then they concluded that imputation methods based on machine learning algorithms are more fit than statistical techniques.

Han C. W. Hsiao et al. [10] analyzed the risk by utilizing a specified area in Taichung, Taiwan, based on four specific categories of cardiovascular disease with respect to factors such as gender, age, region, and month using deep learning approach.

In this paper, to overcome these difficulties, some measures and three different supervised learning algorithms were used for developing better risk prediction models.

41.3 Proposed System

The proposed system includes some steps that can be shown using architecture.

Figure 41.1 shows the architecture of the proposed system. The major steps involved in the proposed system are: (1) The heart disease data from UCI repository is collected. (2) After that, preprocessing should be handled for cleaning the dataset (if there are any missing entries or noisy data). In this work, missing entries are filled using the KNN classifier. (3) Then classification should be performed using Evolutionary Data-Concise Artificial Immune Recognition System (EDC-AIRS), Support Vector Machines (SVM), and Decision Tree (DT). (4) After that, based on their performance, validation measures will be calculated. It shows the best classifier for developing the risk prediction model for heart disease dataset among the three classifiers. Both decision tree and SVM work well on the heart disease dataset. In Sect. 41.3.1, details of heart disease dataset are provided. In Sects. 41.3.2 and 41.3.3, preprocessing and classification using machine learning algorithms are discussed.

41.3.1 Dataset

The dataset used in this paper are depicted in Table 41.1.

This dataset consists of 14 attributes, namely, age, sex, chest pain (cp), trestbps, chol, FBS, restecg, thalach, exang, oldpeak, slope, CA, thal, and class. The chest pain types are broadly classified into four types, namely, typical angina, atypical angina, nonanginal pain, and asymptomatic, according to the UCI repository heart disease dataset.

Fig. 41.1 Architecture of the proposed system

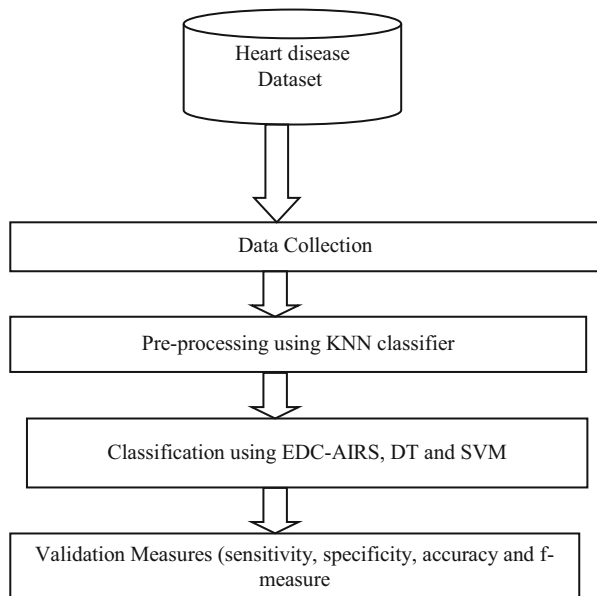


Table 41.1 Dataset

Dataset	Instances	Attributes
Heart data	303	14

Typical angina is the discomfort that is noted when the heart does not get enough blood or oxygen. This is caused by blockage or plaque buildup in the coronary arteries. Atypical chest pain describes discomfort or pain centered in the chest that is not cardiac pain, chest pain not heart related and not of burning quality.

FBS refers to the fasting blood sugar level. Generally if the blood sugar level is less than 100 mg/dL, then the patient is normal. If FBS is between 100 and 125 mg/dL, then it can be considered as prediabetes. If it is 126 mg/dL or higher than that, the patient has diabetes [11]. For predicting CVD, blood sugar level test is considered as one of the main attributes.

An ECG records the electrical activity of our heart at rest. It provides information about our heart rate and rhythm and shows if there is enlargement of the heart due to high blood pressure (hypertension) or evidence of a previous heart attack (myocardial infarction).

The class attribute is the resultant attribute which consists of five classes which indicate the presence and absence of heart disease. Further, these classes provide the risk status level in the form of low risk, medium risk, and high risk. These three classes come under class 1 which indicates the presence of disease and class 0 which indicates the absence. However this dataset consists of a significant percentage of missing data and a highly skewed data distribution. Hence, for effective analysis, preprocessing is performed.

41.3.2 Preprocessing

Incomplete, inconsistent, and noisy data are commonly found in large real-world databases. Therefore data preprocessing is very essential which can be achieved by data cleaning, data integration, data transformation, and data reduction. It is essential to obtain the pure data.

Preprocessing is the process of substituting missing entries in a dataset with possible values and aims to improve the quality of the data. Data cleaning will work to clean the data by filling in the missing values [12]. In this paper, it was performed using K-nearest neighbor (KNN) because of its excellent performance in estimating in missing values [6, 9]. Moreover, it has the capability to estimate both qualitative and quantitative attributes.

41.3.3 Classification

Classification is a data mining technique that assigns categories to a collection of data in order to aid in more accurate predictions and analysis. It is the process of finding a model that describes and distinguishes data classes for the purpose of being able to use the model to predict the class of objects whose class label is unknown [13]. The derived model is based on the analysis of a set of training data.

Classification techniques are of various types. In this paper, three different supervised learning classification algorithms, namely, EDC-AIRS, Decision Tree (DT), and SVM, have been used.

41.3.4 EDC-AIRS

Artificial Immune Recognition System (AIRS) algorithm offers a promising methodology for data classification. It is an immune-inspired supervised learning algorithm that works efficiently and has shown comparable performance with respect to other classifier algorithms.

EDC-AIRS algorithm is a supervised classification algorithm inspired by the principles and processes associated with the human immune system. It performs grouping by first building a pool of memory cells that are representative of the training data through repetitive optimization of the memory cells. For each training instance, it first undergoes the memory cell identification stage where its affinity with the memory cell in MCP is computed. EDC-AIRS has the ability to robustly adapt to the different density, distribution and characteristics exhibited by each data class. This provides competitive advantages for the algorithm to better characterize and learn the underlying pattern of the data. It employs several natural immune mechanisms. This immune system possesses properties such as the capability of recognition, memory acquisition, diversity, and self-regulation, making it highly suitable for learning patterns that underlie a data. Although AIRS algorithm has shown to be an effective classification algorithm, some useful immune mechanisms are yet to be exploited by the algorithm.

Optimization was carried out by robustly adapting the memory cells to the different density, distribution, and characteristics exhibited by each data class in the training data. Finally, with the utilization of the produced memory cells pool, KNN is used to classify unseen data observations. It works well on large datasets. This algorithm, when tested on several widely benchmark datasets, has exhibited greatly aggressive grouping execution [7].

41.3.5 Decision Tree (DT)

Decision Tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes, from a given dataset. Generally the set of records available for developing classification methods is divided into two disjoint subsets such as training and test set. The training set is used to derive the classifier, while the test set is used to measure the accuracy of the classifier. The accuracy of the classifier is determined by the percentage of the test samples that are correctly classified.

The major strengths of the decision tree methods are they can handle both numerical and categorical attributes. They can provide a clear indication of which fields are most important for prediction or classification [14]. Decision tree consists of nodes and branches that connect the nodes. The nodes located at the bottom of the tree are called leaves and indicate classes. The top node in the tree except the leaves is called root, and it contains all training examples that are to be divided into classes. All nodes except the leaves are called decision nodes, since they specify decision to be performed at each node based on a single feature. Each decision node has a number of children nodes, equal to the number of values that a given feature assumed. The uncertainty in each node is reduced by choosing the feature which reduces its entropy. To achieve this result, information gain (IG) and Gini Index (GI) can be used.

In this work, information gain is used. Information gain measures expected reduction in entropy caused by the value of a feature F_j . It is used to select the best feature at each step of growing a decision tree.

$$\text{Information Gain}(S, F_i) = \text{Entropy}(S) - \sum_{v_i \in V_{F_j}} |S_{v_i}|/|S| \cdot \text{Entropy}(S_{v_i})$$

where V_{F_j} is a set of all possible values of feature F_j and S_{v_i} is a subset of S , for which feature F_j has value v_i . In this paper, the heart disease data attributes such as thal, oldpeak, thalach, and chest pain(cp) were randomly chosen for developing a decision tree for predicting the risk status of cardiovascular disease. According to the results of these four attributes, the patients can be placed on any one of the five classes. These attributes can vary on different training and test datasets.

These four attributes can be changed randomly by the variation in the training and testing data.

Figures 41.2 and 41.3 show the different sample decision trees that can be changed on different training and test datasets. Through this classifier, the importance of the particular attributes of heart disease in predicting the risk can also be found.

41.3.6 Support Vector Machine (SVM)

SVM is a supervised learning algorithm that is capable of yielding generalization performance on an extensive area of problems. It is a learning machine that can

Fig. 41.2 Sample tree1

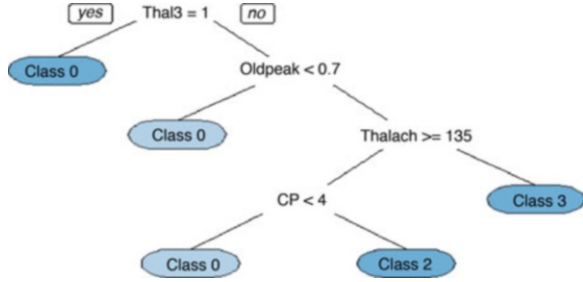
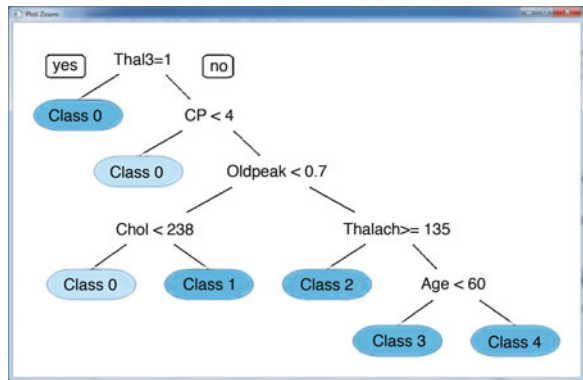


Fig. 41.3 Sample tree2



perform binary classification and regression estimation tasks. SVM minimize the expected error rather than minimizing the classification error. SVM incorporate structured risk minimization into the classification. It solves both linearly and nonlinearly separable problems [8].

In a supervised classification, there will be a training set and a test set. The classification error is calculated based on the ratio of misclassified test data items to the whole data set. However, the classification scheme is generated from the training data set. SVM create a classifier with minimized expected probability of error.

SVM mainly focuses on determining two maximally apart hyperplanes. It performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two or more categories, whereby the margin of separation between the different categories is maximized. If the two classes are nonseparable, we can still look for the hyperplane that maximizes the margin and that minimizes a quantity proportional to the number of misclassification errors.

The key idea of SVM is that the classifier must reduce the expected classification error and the confidence interval to ensure good generalization; that is, the wider the gap between the planes, the smaller the generalization error [15]. The hyperplane constructed in SVM is very high or infinite dimensional area which might be used for classification, regression or alternative tasks like outliers detection.

In this paper, after preprocessing, SVM algorithm should be implemented for heart disease dataset using R tool with 0.75 as training data and remaining 0.25 as testing data. SVM performs better on noisy data than EDC-AIRS.

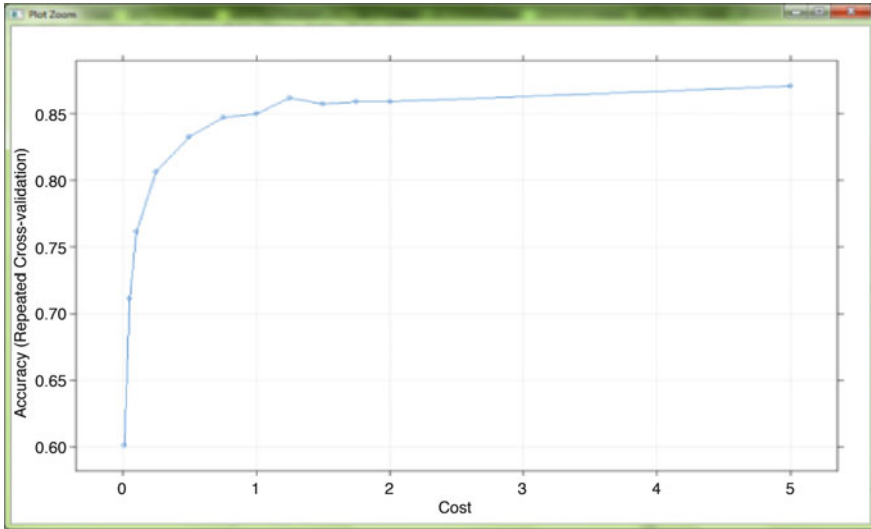


Fig. 41.4 SVM cost graph

Figure 41.4 represents the cost graph of SVM. It plots the accuracy of SVM by tuning the kernel parameter. To improve the accuracy in SVM, parameters involved in SVM should be tuned.

41.4 Experimental Results

This work is implemented using R tool and visualization of results using tableau tool. Table 41.1 provides the details of the first 6 records of the heart disease dataset with 5 attributes among 14 attributes after the removal of missing entries using KNN. It is a sample table of heart disease dataset which consists of 303 instances and 14 attributes.

Table 41.2 shows some sample records of the heart disease dataset after using KNN. KNN is used to fill the missing entries of the dataset under the preprocessing method. Among other classifiers, KNN is very efficient in the data cleaning process. The first six attributes represent age, sex, chest pain type (cp), resting blood pressure (restbtps) which is one of the types of heart disease indicator, serum cholesterol level (chol), and thalach. Serum is a clear liquid that can be separated from clotted blood. It differs from plasma, the liquid portion of normal unclotted blood containing the red and white cells and platelets. Thalach indicates the maximum heart rate achieved for the patients.

In this paper, the heart disease dataset were applied on three classifiers, namely, EDC-AIRS, DT, and SVM. Among the three classifiers, SVM provides better result in terms of validation measures such as accuracy, sensitivity, specificity, and F-measure. Table 41.2 presents the accuracy of three classifiers.

Table 41.2 Sample records of heart disease dataset after using KNN

Age	Sex	CP	Trestbps	Chol	Thalach
63	1	1	145	233	150
37	1	3	130	250	187
41	0	2	130	204	172
56	1	2	120	236	178
57	0	4	120	354	163
57	1	4	140	192	148

Table 41.3 Comparison of accuracy

Algorithms	Accuracy (in %)
EDC-AIRS	56.76
DT	62.16
SVM	85.14

Table 41.4 Comparison of validation measures

Algorithms	Sensitivity	Specificity	F-Measure
EDC-AIRS	0.36	0.88	0.51
DT	0.33	0.86	0.48
SVM	0.65	0.97	0.78

Table 41.3 presents the validation measures in terms of sensitivity, specificity, and F-measure of three classifiers. Sensitivity is the ratio of correctly predicted positive observations to the total predicted positive observations that can be calculated through

$$\text{Sensitivity} = TP / (TP + FP)$$

Specificity is the ratio of correctly predicted positive observations to the all observations in actual class. That can be

$$\text{Specificity} = TP / (TP + FP)$$

F-measure is given as the weighted average of sensitivity and specificity which can be

$$F - \text{measure} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

These three validation measures calculate the efficiency of the algorithms. It will vary on different datasets. Further, McNemar’s test can also be conducted. It first records the prediction outcomes by each algorithm and tested using the validation dataset. In this paper, McNemar’s test is not conducted (Table 41.4).

Graphical representations of accuracy and validation measures of each classifier are as follows:

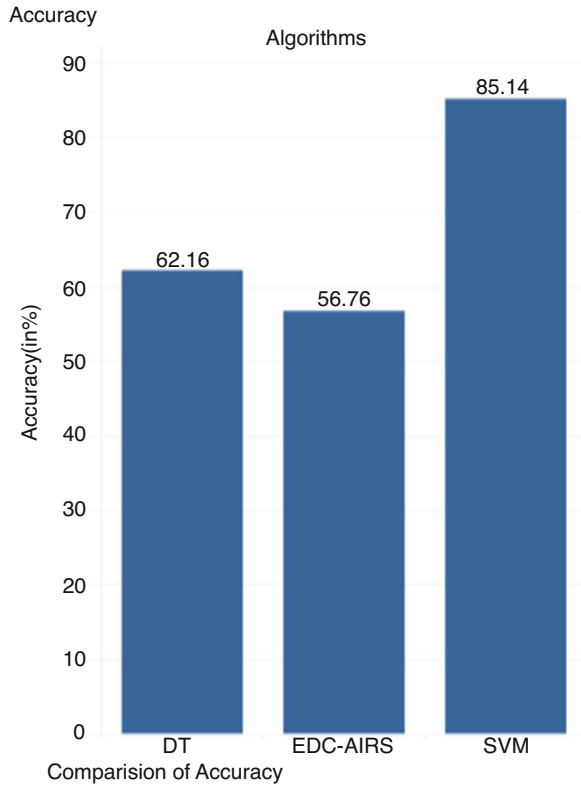
Fig. 41.5 Performance analysis

Figure 41.5 represents the performance analysis through accuracy of the three classifiers EDC-AIRS, DT, and SVM. Figure 41.6 represents the comparison of validation measures of the three classifiers. According to this paper, the result indicates that the Support Vector Machine classifier is more suitable than EDC-AIRS and Decision Tree for the heart disease dataset. SVM suits more for noisy data than EDC-AIRS.

41.5 Conclusion

CVD is one of the most common causes of death worldwide and represents a major financial burden for national economies. The epidemic of CVD has caused huge losses and caught the attention of society. 90% of CVD is preventable. Globally 17.9 million deaths due to CVD occurred in 2015 [16]. Early detection of cardiovascular disease risk is very important clinically. Its main goal is to better meet the patient's needs and clinician's expectation. In this paper, risk prediction models developed using classification techniques were analyzed. From the obtained accuracy and

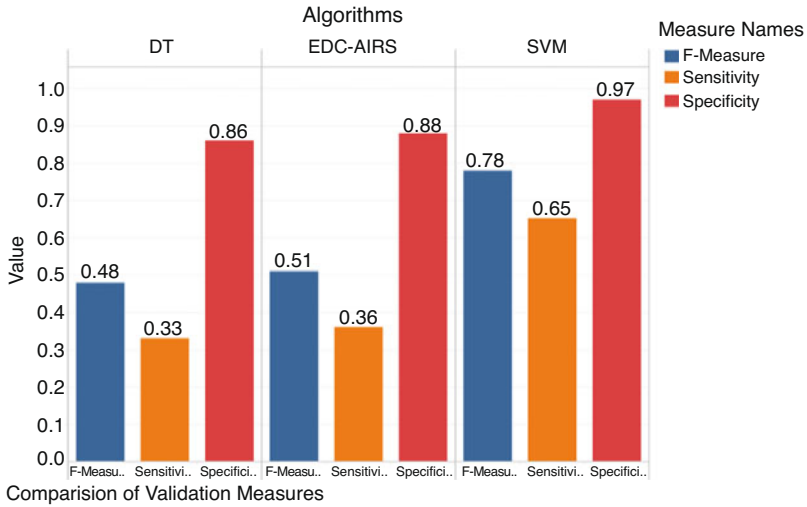


Fig. 41.6 Result of validation measures

validation measures, it is concluded that SVM is more suitable than EDC-AIRS and DT for the heart disease dataset. SVM performs better on noisy data than EDC-AIRS.

Further, ensemble techniques such as bagging, boosting and advanced feature engineering, and artificial intelligence can be applied for much more refined and progressive risk prediction without compromising accuracy.

References

1. Wilson PW, D'agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB (1998) Prediction of coronary heart disease using risk factor categories. *Circ Cardiovasc Genet* 97(18):1837–1847
2. Abbott R, Curb J, Rodriguez B, Masaki K, Yano K, Schatz I, Ross G, Petrovitch H (2002) Age-related changes in risk factor effects on the incidence of coronary heart disease. *Ann Epidemiol* 12(3):173–181
3. Sen SK (2017) Predicting and diagnosing of heart disease using machine learning algorithms. *Int J Eng Comput Sci* 6(6):21623–21631. ISSN:2319-7242
4. Osuna E, Freund R, Girosit F (1997) Training support vector machines: an application to face detection. In: *Proceedings of Computer Vision and Pattern Recognition*, pp 130–136
5. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman R (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17(6):520–525
6. Tay D, Poh CL, Van Reeth E, Kitney RI (2015) The effect of Sample age and prediction resolution on myocardial infarction risk prediction. *IEEE J Biomed Health Inform* 19(3):1178–1185

7. Bia D, Zocalo Y, Farro I, Torrado J, Florio L, Lluberas R, Armentano R (2012) Health informatics design for assisted diagnosis of sub-clinical atherosclerosis, structural and functional arterial age calculus and patient-specific cardiovascular risk evaluation. *IEEE Trans Inf Technol Biomed* 16(5):943–951
8. Hsiao HCW, Chen SHF, Tsai JJP (2016) Deep learning for risk analysis of specific cardiovascular diseases using environmental data and outpatient records. In: *IEEE 16th international conference on Bioinformatics and Bioengineering (BIBE)*. ISSN:2471-7819, pp 369–372
9. Tay D, Poh C, Kitney R (2013) An evolutionary data-conscious artificial immune recognition system. In: *Proceedings of 15th annual conference on Genetic and Evolutionary Computation Conference*, pp 1101–1108
10. Song X, Mitnitski A, Cox J, Rockwood K (2004) Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Medinfo* 107(1):736–740
11. www.mayoclinic.org
12. Jerez J, Molina I, Garcia-Laencina P, Alba E, Ribelles N, Martn M, Franco L (2010) Missing data imputation using statistical and machine learning methods in a real breast cancer problem. *Artif Intell Med* 50(2):105–115
13. Suresh Kumar TV, Eswara Reddy B, Kallimani JS. *Data mining principles and applications*. Elsevier, A Division of Reed Elsevier India Private Limited, Amsterdam
14. Han J, Kambar M *Data mining concepts and techniques*. Morgan Kaufmann Publishers, Boston
15. Pujari AK. *Data mining techniques*. Universities Press (India) Private Limited, Hyderabad
16. Phan JH, Quo C, Wang M (2012) Cardiovascular genomics: a biomarker identification pipeline. *IEEE Trans Inf Technol Biomed* 16(5):809–822

Chapter 42

Safest Secure and Consistent Data Services in the Storage of Cloud Computing



Geethu Mary George and L. S. Jayashree

Abstract Cloud computing is the greatest learning in the computing field and a dreamed vision of computing as a utility so to enjoy the on-demand high-quality applications. Cloud security is the critical factor that places an imperative role in maintaining the secure and reliable data services. In large-range cloud computing, a large pool of erasable, usable, and accessible virtualized resources are used as hardware development platforms and/or sources. These resources can be vigorously reconfigured to adjust a variable load allowing also for optimum resource utilization. The pool of resource is typically exploited by a peer-to-peer use model in which guarantees are presented by the infrastructure provided by means of customized service-level architecture (SLA). The hierarchical structure has been proven effective for solving data storage issues as well as data integrity by giving data protection during the full life span. Cloud computing is related to numerous technologies, and the convergence of diverse technologies has emerged to be called cloud computing. Storage in the cloud provides attractive cost and high-quality applications on large data storage. Security offerings and capability continue to increase and vary between cloud providers. Cloud offers greater convenience to users toward data because they will not bother about the direct hardware management. For security issues, a secret key is generated. Key consideration is to efficiently detect any unauthorized data corruption and modification which arises due to byzantine failures. Cloud service providers (CSP) are separate administrative entities, where data outsourcing is actually relinquishing user's ultimate control over the fate of their data. As an outcome, the accuracy of the data in the cloud is being set at a high risk. In distributed cloud servers, all these inconsistencies are detected and data is guaranteed. The main proposed objective of this chapter is to develop an auditing mechanism with a homomorphic token key for security purposes. By using this secret token, we will easily be able to locate errors and also the root cause of the error. By the error recovery algorithm, we recover these corrupted files and locate the error.

G. M. George (✉) · L. S. Jayashree
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

Keywords Cloud computing · Data storage security · Error localization · Data integrity · Token generation · Pseudorandom data

Abbreviations

SLA	Service-level architecture
CSP	Cloud service provider
SaaS	Software as a service
PaaS	Platform as a service
IaaS	Infrastructure as a service
SLA	Service-level agreement
TPA	Third-party auditor
CS	Cloud server
GF	Galois field
SRF	Sobol random function
SRP	Sobol random permutation
MAC	Machine authenticated code

42.1 Introduction

Cloud computing is the evolution and convergence of a number of mature and fast-maturing technology market threads such as virtualization, utility effectiveness computing, and software-like service. The services of cloud defines three delivery models, namely, software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS). It also defines four deployment models, namely, private, public, community, and hybrid. Cloud computing already extends beyond the sum of these three to represent the core of what is fast becoming the most disruptive computing model. Today, hosts that are connected to the Internet include mainly servers, client computers, etc. The services of cloud include three delivery models, namely, software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS). It also defines four deployment models, namely, private, public, community, and hybrid. Some of the chief firms like Amazon, Microsoft, and Google enfold, implemented the “CLOUD,” and have been using it to speed up their business. However, the Internet and network configurations are meeting the growing danger of security breaches. Moreover, an end host can easily join the network and can easily communicate with any other host by exchanging packets within the network or outside the network. Thus openness and scalability are being the encouraging features of the Internet. Cloud computing moves databases

and application software to large data spaces, where managing of data and services may not be fully trustworthy. Cloud data storage security has been a vital phase of quality of service.

Cloud technology is adapted by large companies as a way of streamlining their IT infrastructures; it is currently being embraced by increasing numbers of smaller businesses around the country. Enhancing the cloud assures minor cost, small companies are judging that cloud computing can also tender improved flexibility, enhanced security, and poorer risks. In its simplest form, cloud computing includes a business handing more responsibility meant for its IT systems to a third-party service provider. Rather than needing to worry about running supercomputer servers, updating software, and performing data back-ups, the business “rents” the capability it needs, accessing applications with data so as to sit on remote servers through the Internet. From users’ perspective, it includes both individuals and IT centers. As a disruptive technology with insightful implications, cloud computing is transferring the extreme nature of how businesses use information technology. One fundamental aspect of these prices, storing data remotely to the cloud in a flexible on-demand manner brings appealing benefits: relief of the burden for storage management, universal data access with location independence, and evading of wealthy expenditure on hardware, software, furthermore personnel maintenance, etc. Since cloud service providers are detached from administrative entities, data outsourcing is really relinquishing user’s ultimate control above the fate of their data. As an effect, the exactness of the data in the cloud is worldly put at a risk due to the subsequent reasons. Firstly, although the infrastructures that lower the cloud are more powerful and reliable than personal computing devices, they are unmoving facing the broad range of both internal and external threats for records reliability. Secondly, there do exist different motivations for CSP to act unfaithfully toward the cloud users concerning their outsourced data status. These pack attacks are simple to set up, difficult to stop and control, and very efficient. There are various types of such attacks. Some groups separate attacks into three categories: bandwidth attacks, attacks based on communication protocols where the flaws in the protocol are found out and attacks are carried out, and logic attacks which uses various logics.

Timeouts may occur, which causes retransmission, generating even more traffic in the network. An attacker can consume bandwidth by creating any traffic by flooding the packets at all on a network connection. These attacks may undergo security weakness in the operating systems of attached computers as well as vulnerabilities in the Internet routers and other network devices. The security problem will have an outcome on the use of Internet next to the density and complexity of protocols and applications.

For this purpose, confidentiality can be protected by encrypting message plus token with either the receiver’s public key or shared secret key. To ensure the privacy of user data, an effective and audible distributed scheme with two silent features is presented, opposing its predecessors.

42.2 Related Works

Many computing-based counteract measures have been projected to address the problem of security issues. Cloud computing inevitably poses new challenging security threats for a number of reasons. First, cryptographic primitives for the purpose of data security protection cannot be adopted due to the users' loss control of data under cloud computing. Second, cloud computing is not a third-party data storehouse. The stored records in the cloud could be frequently revised with the users, counting operations like insertion, removal, envisioning, affixing, reordering, etc.

Mehaul A. Shah [1] proposed that third-party auditing is important in creating an online service-oriented economy to evaluate risks and increases the efficiency of insurance-based risk mitigation. The effects of adversary's different strategies for launching blind attacks are also analyzed separately. The approaches and system hooks that support both internal and external audit services explain motivations for service providers and auditors to adopt these approaches and list challenges that need to be solved for auditing to become a reality.

Third-party auditing is an accepted method for connecting a relation between two parties with potentially different incentives. One way is to rely on a trusted third-party auditor (TPA), who has sufficient access to the providers' environment. An auditor needs to understand the agreement between a customer and a provider and checks up to which a provider might not meet the agreement. The agreement is termed as service-level agreement (SLA). The SLA for storage services can include data integrity, security, data outs, and privacy. An auditor performs checks for process adherence and service quality. They perform these checks using well-defined interfaces. Internal audits evaluate the structure and process within a service to ensure that services can follow the best practices to meet its objectives.

Wenjing Lou [2] proposed a cloud data storage security, which is the main important aspect of storage service. To verify the correctness of user data in cloud, we can use an effective and flexible distributed scheme. First, by utilizing the homomorphic token with distributed verification of erasure-coded data, this method achieves the integration of storage correctness insurance and identification of misbehaving servers. Secondly, it supports secure and efficient dynamic operation on data blocks, which includes data bring up to date, erase and affix. Extensive security and performance analysis shows that these methods are highly efficient and resilient against Byzantine failure and even from colluding server attacks. Unlike most prior works for remote data integrity, the new scheme support on data secure and efficient dynamic operations across distributed servers, the challenge-response protocol provides the localization of data error.

External audits evaluate the quality of service through external interfaces. External audits can confirm only past behavior, so without internal audits, we can predict problems or risk exposure. The ideal goals of auditing services in storage include establishing standards for comparison, minimizing auditing cost, protecting customer data privacy, avoiding prescribing technology, and protecting proprietary

provider information. To ensure the security and dependability for cloud storage, aim to design an efficient technique for dynamic verification of data and achieve storage correctness, dynamic data support to maintain the same level of storage correctness assurance, dependability to minimize the effect caused by the data errors, and lightweight to enable users to perform checks with minimum overhead.

Smitha Sundareswaran [3] proposed a major feature of cloud services that the users data are processed in an unknown machines that users do not own or operate. To overcome this problem, a novel highly decentralized information accountability framework to keep track of actual usage of data in the cloud. An object-oriented approach enables enclosing our sorting mechanism together with users data and policies. The proposed novel approach, namely, cloud information accountability (CIA), is based on the notion of information accountability. One of the main features of cloud information accountability lies in its ability to maintain lightweight and powerful responsibility that includes access control, usage control, and authentication. In association with the accountability feature, two distinct modes have been developed: push mode and pull mode. The push mode explains logs being sporadically sent to the data owner, whereas the pull mode retrieves the logs as needed. Users will send data along with any polices such as access control policies and logging policies that they want to impose into JAR files to cloud service providers. JAR files control and extend the programmable ability of files to automatically log the procedure of users data by an entity in the cloud. The logging mechanism enforcement termed as “strong binding” polices travel with the data. This strong binding exists even when copies of JAR are created; thus users will have control over the whole data at any location. Such decentralized sorting mechanism meets the vibrant nature of cloud, however also challenges in ensuring the integrity of logging.

Hsiao-Ying Lin et al. [4] proposed a cloud storage space system, which consists of a collection of storage servers and provides longstanding storage services in excess of the Internet. Storing data within a third party’s cloud structure causes stern concern above data confidentiality. General encryption schemes guard data confidentiality but also border the functionality of the storage system because a few operations are supported over encrypted data. To ensure that the communication is secure, a challenge server is deployed for the purpose of issuing keys, and its main aim is assuring the number of clients connected with the server and synchronizing the client’s with the server as well. Thus protection of the challenge server is quite important, and defending against attacks to the server is also necessary. The paper cites a threshold proxy re-encryption scheme and incorporates it with a decentralized erasure code such that a secure distributed storage system is formulated. This distributed storage system does not merely support security by way of robust data storage and retrieval; however, it also lets a user frontward his data in the storage servers to an additional user devoid of retrieving the data back. A decentralized erasure code is appropriate for use in a distributed storage scheme. After the message symbols are sent to storage servers, all storage servers independently computes a code word symbol for the received message symbols and stores it. This completes the encoding and storing process.

Vrushali W. Basatwar [5]. proposed a flexible distributed storage veracity auditing method, utilizing the homomorphic token along with distributed erasure-coded data. Our method achieves the integrity of storage correctness guaranty and identification of misbehaving servers. Whenever data modifications or deletions are detected during the storage, correctness verification and error localization across cloud servers are done. This scheme efficiently detects data corruptions and achieves the guaranty of file retrievability.

Shacham et al. [6] introduced a new model of POR, which enables an unrestricted number of queries for public verifiability with less overhead. This scheme achieves the guarantee of data availability, reliability, and integrity. However, these schemes also provide full protection to user data in cloud computing, since pseudorandom data possibly will not be able to cover the entire data.

To cope with this, Abdinandan P [1] also proposed a solution for enabling public audit ability for cloud data storage security, which is of critical importance, so that users can choose an external audit party to check the reliability of outsourced data as soon as needed. To securely set up an efficient third-party auditor (TPA), the following two fundamental requirements have to be met: First, the TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data and initiate no additional online burden to the cloud user. Second, the third-party auditing process should bring in no new vulnerabilities toward client data isolation. In this work, we utilize and uniquely combine the public key-based homomorphic authenticator with a random mask to achieve the privacy-preserving public cloud data audit system, which meets all the above necessities. To maintain efficient handling of multiple auditing tasks, we additionally explore the practice of a bilinear cumulative signature to enlarge our main result into a multi-user location, where TPA can perform multiple auditing tasks simultaneously. These works explain cloud data storage service involving three different entities: the cloud user, who has a large amount of data files to be stored in the cloud; the cloud server (CS), which is controlled by the cloud service provider (CSP) to make available data storage space service and has momentous storage space and computation resources; and the third-party auditor (TPA), who has expertise and capabilities that cloud users do not have and is trusted to assess the cloud storage service security on behalf of the user ahead request. Users rely on the cloud server for cloud data storage and maintenance. They may also dynamically interrelate with the cloud server that makes use of and updates their stored data for a combination of application purposes. The users may route to TPA for ensuring the storage space security of their outsourced data while hoping to maintain their data private opening third-party auditor (TPA). The TPA, who is the business of auditing, is consistent and independent and thus has no incentive to plot with either the cloud server or the users during the auditing process. Third party should be able to efficiently audit the cloud data storage with no local copy of data and without bringing in additional online burden to cloud users.

Kennadi D et al. [7] proposed a hypothetical structure for the design of Proof of Retrievability. It improves the JK [8] and SW [6] models. All the schemes produce feeble security, since they work merely for single server. Recently, Wang et al. [1]

describes a homomorphic circulated verification scheme using pseudorandom data to verify the storage correctness of user data in cloud. This method achieves the deposit of data availability, reliability, and integrity. However, this scheme moreover did not provide a full shield to user data in cloud computing because pseudorandom data would not cover up the entire data. Sobol sequence [4] is an example of quasi-random low-discrepancy sequences.

Cong Wang [6] proposed an outsourcing storage which is of low cost and complexity of lasting large data storage. This service can be used for avoiding overall control of data owners. A large amount of cloud data and owners computing capabilities makes the task of data correctness auditing correctness inexpensive in a cloud environment and formidable for individuals. Third-party auditor gains knowledge of unauthorized information through auditing process especially from data owners entailed in cloud. To avoid these sensible services, secure cloud storage is to be maintained for correctness assurance even though data is dynamically changing. Data made by enterprise well individuals are stored in complex data management systems for its flexibility and cost-effective savings. Challenging problems is related to a huge amount of outsourced data and a large number of on-demand data users, as it is extremely complicated to meet the requirements of performance, scalability, usability, and easy understanding.

42.2.1 Problem and System Model Definitions

The problem states network representative architecture for cloud data storage, which contains three parts as shown in Fig. 42.1, including users, cloud service provider (CSP), and third-party auditor (TPA).

- (1) **User:** Users, which stores data in the cloud and relies on the cloud for data storage space and computation, can be either an individual or a group.
- (2) **Cloud Service Provider (CSP):** A CSP, who has major resources and expertise in building and running distributed cloud storage servers, owns and operates subsist cloud computing systems.
- (3) **Third-Party Auditor:** An optional TPA, who has knowledge and capability that users may not have, is trusted to assess and depict the risk of cloud storage services on behalf of the users upon request. The main process is when a user needs to check the data integrity and data dynamics, Third-party auditor (TPA) checks those are valid for the verification process by getting a query from the user then retrieves the corresponding key and verifies the data dynamics from cloud storage.

Figure 42.1 explains the cloud architecture which explains from the beginning to the bottom end and also how operations are carried over and what methodical procedures are carried over [9]. The main is TPA, which starts its operation from user registration till all data are stored in cloud storage.

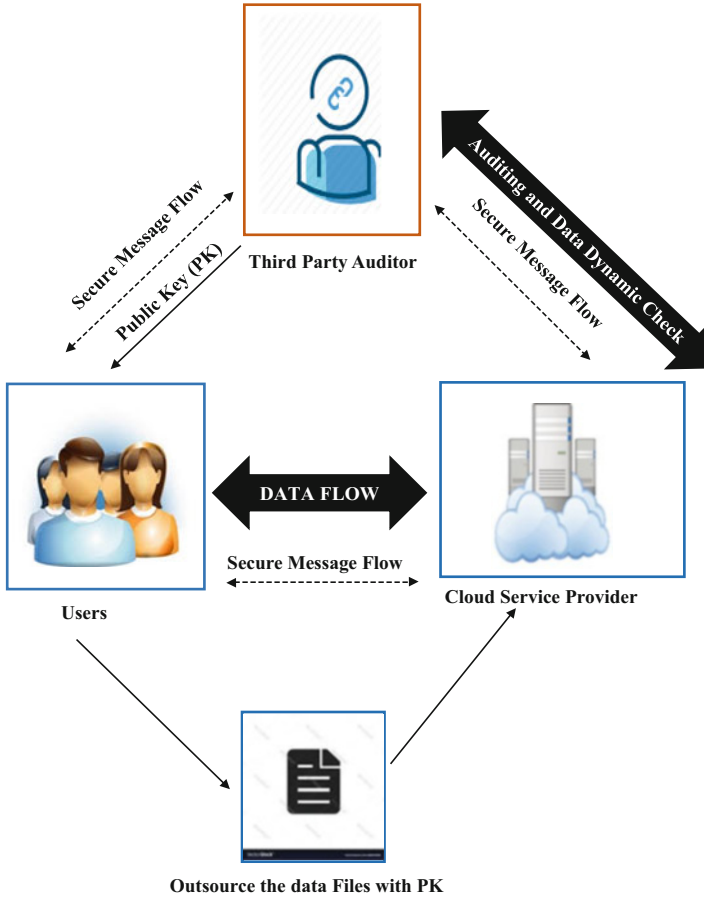


Fig. 42.1 Cloud storage architecture

42.3 Overview

Cloud computing had envisioned the next generation information technology (IT) architecture for enterprises, due to its lengthy list of unprecedented advantages in the IT history: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid resource elasticity, and transference of risk paradigm shifting indicate that data are being centralized or outsourced to the cloud. Different methods where information stored in cloud can undergo security issues that may include network sniffing – receivers data that is not encrypted as well as transferred securely may affect the network and lead to network sniffing [10]. The most severe attack is authentication attack which can be oppressed to an attacker based on the authentication method.

The design allows users to audit the cloud storage with two approaches: one with very lightweight communication and the other based on computation cost [11].

Lightweight communication enables users to perform storage correctness checks with minimum overhead in order to save time, computation resources, and even the associated online burden of users; it also supplies the extension to support third-party auditing, wherever users can carefully delegate the integrity checking task [12, 13]. Users can perform normal authentication fastly and conveniently. It was as well suggested to use nonuniform nodule distribution to mitigate message relay [14]. The first two approaches have limited effectiveness since storage around server is very likely to be critical to server connectivity and cannot be skipped, while the third approach reduces cost and complexity, which is the functional basis of any cloud computing. After user registration with third-party auditor (TPA), it provides a security key and is used for further transactions. The user performs both validations and verifications in the cloud server. The third-party auditor is the most trusted party who exposes the hazards and risk of cloud storage space service upon user request.

42.3.1 Design Objective

To ensure security and reliability for cloud data storage, an adversary model is developed to provide additional security to users data stored in cloud computing, i.e., guarantee the availability, dependability, and integrity of data [15]. Users also have to perform storage correctness checks with minimum overhead [14, 16]. The objective includes (1) public assessment to allow the TPA to make sure the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing; (2) cloud storage correctness to make sure that the user data are stored properly and kept intact all the time in cloud; (3) lightweight to allow users and TPA to perform auditing with least computation and communication overhead; and (4) fast localization of data error to effectively and successfully place the malfunctioning server when data corruption has been detected.

42.3.2 Preliminaries and Notations (Table 42.1)

The three main methods incorporated to illustrate how these techniques support for TPA in lead delegations from multiple users include [17]:

- Homomorphic token generation
- Record verification and error localization
- File repossession and error revival

Homomorphic Token Precomputation

In cloud data storage, this method relies on the technique to distribute the entire data file across a set of $k = m + n$ cloud servers for data availability and reliability. A $(m + n, n)$ Reed-Solomon erasure coding can be used to generate n parity blocks from

Table 42.1 Key parameters

<i>Df</i>	The data file to be stored, to <i>Df</i> can be denoted as a matrix of <i>m</i> equal-sized data vectors, each consisting of <i>l</i> block; all these data blocks belongs to Galois Field $GF(2^p)$ where $p = 8$ or 32
<i>A</i>	The distributed matrix for Reed-Solomon coding
<i>G</i>	The encoded file matrix, which includes a set of $k = m + n$ vectors, each having <i>l</i> block
<i>fkey</i>	Sobol Random Function (SRF), which is defined as $f: \{0,1\}^* \times \text{key} \rightarrow GF(2^p)$
<i>pkey</i>	Sobol Random Permutation (SRP) which is defined as $\hat{E}: \{0,1\}^{\log_2(l)} \times \text{key} \rightarrow \{0,1\}^{\log_2(l)}$
<i>ver</i>	Records the times the blocks have been modified. Initially we assume <i>ver</i> is 0 for all data blocks
<i>sijver</i>	The seed for SRF, which depends on the file name, block index <i>i</i> , the server position <i>j</i> , and version number <i>ver</i>

m data blocks in such a way that the original data file can be reconstructed from any *m* out of *m + n*, i.e., the original file can survive from any failures of *m + n* servers without any data loss and with less space overhead [18, 19]. The main design is detailed as follows: before file sharing, the user precomputes a firm quantity of short verification tokens on individual vector, each token wrapper a random subset of data blocks. Later on, when the user desires to make sure the storage correctness for the data in the cloud, it challenges the cloud servers with a set of randomly generated block index [20].

The cloud server will obtain a reply message by executing token file *F* accurately at the time of audit and its verification metadata as inputs. When in receipt of challenge, each cloud server computes a short “signature” over the particular blocks and proceeds them to the user [21]. The standards of these signatures must match the comparable tokens precomputed by the user. After token generation, the user has the option of either keeping the precomputed tokens close by or storing them in encrypted form on the cloud servers. To create a situation of *m* token from server *k*, the user acts as follows:

Steps Undertaken

- (1) Developed a typical *s* data vector with *l* equality vectors in dispersed servers by *fkey*(*i*) and a challenge key by way of master permutation key based on MPRP.
- (2) Computed the set of *v* arbitrarily – select indices in such a way that the original data file can be reconstructed from any *m* out of *m + n*, such that the original file can survive from any failures of *m + n* servers without any data loss and with less space overhead.
- (3) Calculated the token as:
 - (a) On registering with the user, a secret token is generated and includes some necessary cryptographic background for generating a token.
 - (b) The user initializes the public and secret parameters of the system by executing KeyGen and processes the data.
 - (c) If the key parameters match with the user token, verification is complete.

- (4) Pre-computed tokens are elected by the user and subsequently stored locally in an encrypted format on server plane.

The entire servers activate over the similar subset of index, and the requested standards of retorts check the integrity must be accurate codeword resolute by the secret matrix.

Records Certification and Error Localization

Error localization is a key throbs condition for eliminating errors in storage systems. It is also of vital importance to recognize potential threats from external attacks. However, many previous schemes do not clearly consider the crisis of data error localization, thus only providing binary results for the storage verification [22]. This method outperforms those by integrating the misbehaving server identification in our challenge-response protocol: the response values opening from servers for each one challenge not only resolve the correctness of the circulated storage but in addition includes information to place potential data error(s). There are two possible ways to verify the records and situation the actual position of malfunctioning servers by the use of machine authentication code (MAC) to authenticate the data.

Steps Undertaken

- (1) Upload the records blocks in G using their MACs to the server and transmit the correspondent secret key to the TPA.
- (2) TPA retrieves the corrupted blocks with MAC address and verifies the correctness including equality inspection.
- (3) Re-checking over n servers is as follows:
 - (a) $pkey$ to all servers and then the value of T_i has been computed and is a server storing vector.
 - (b) The server includes vector $J(k)$ in D_f and splits those rows précised by permutation key into geometric grouping by assigning the equation functional value from the equation.
 - (c) Upon getting the value from every server, the user takes away unsighted values.

To conclude the user checks the acknowledged value should leave a suitable codeword finds by secret token.

- (4) Geometric combinations of v rows have to be placed in an encoded file matrix. If the value clutches, the challenge is passed.

File Reclamation and Error Revival

The user is competent to restructure the original folder by downloading the data case vectors starting first with m servers, assuming they set the exact response values. By selecting scheme parameters accurately and by continuously conducting verification for more times, it guarantees the successful file recovery with very high probability. When data corruption has been continuously detected, evaluate the precomputed token value, and the retrieved response value can easily find out the misbehaving servers. The user has the right to send back and ask which data has undergone failure

and renovate the correct blocks by erasure correction technique [13]. Each response is computed precisely in the similar way as the token W_i , where both equations have been obtained; hence, the user can note which server is misbehaving by checking the following equation and correcting the recovered files by this core equation.

42.4 Extension to Data Storage Security

The extension to data storage security per server is achieved here. This is because by means of cloud storage, data resides on the web positioned across storage systems than at a designed corporate hosting location. Substituting the token key in, introduced an image for quick actions and making it more reliable, ease of use for data users. This process has less overhead than earlier scheme and supports for building mass updates and provides data accessibility assurance against server failures and supports explicitly data dynamics. Introducing the icon instead of token value generation makes it inexpensive, so savings can be captured as large providers continue to constrain down infrastructure costs. Keyword search is the next proposed work to be implemented to make the users know which data has undergone the process of upload, download, and other dynamic operations and to find out which users have used certain operations.

42.5 Experimental Results

The proposed scheme was analyzed in terms of security along with efficiency. Security analysis focused on the adversary model with generation of homomorphic token.

Figure 42.2 explains third-party auditor key, which gets the time of registration; with this key, it performs further operations. Restarting the machine executes the malicious code.

Figure 42.3 shows the main context where the whole process starts. In some cases, the user may need to perform block-level operations on his data. In this key, search operation is performed with image specification and then in addition generates corresponding values.

Figure 42.4 explains the whole third-party process and its operation. In this module, users can check the data integrity and dynamics; they post a query to the third-party auditor (TPA) and those are validated during the verification process. TPA avoids revealing risks and provides suitable and reasonable solutions to recovered files.

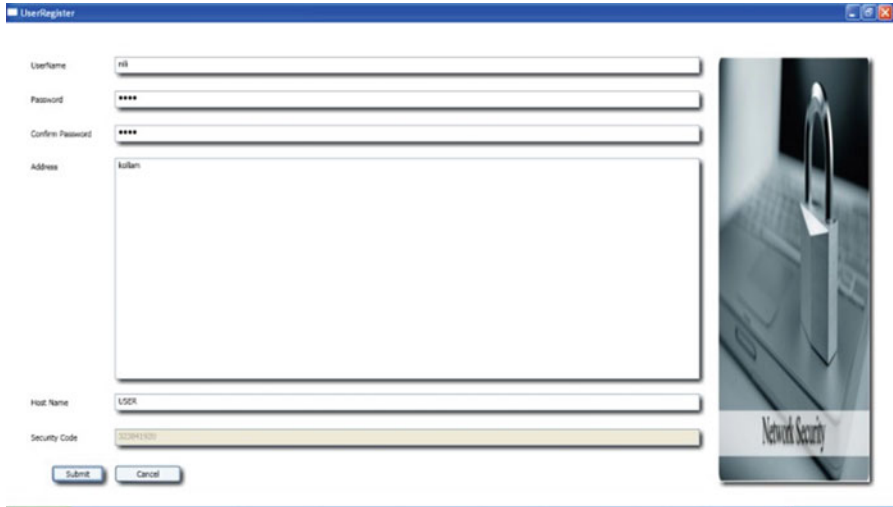


Fig. 42.2 User registration with third-party auditor (TPA) key

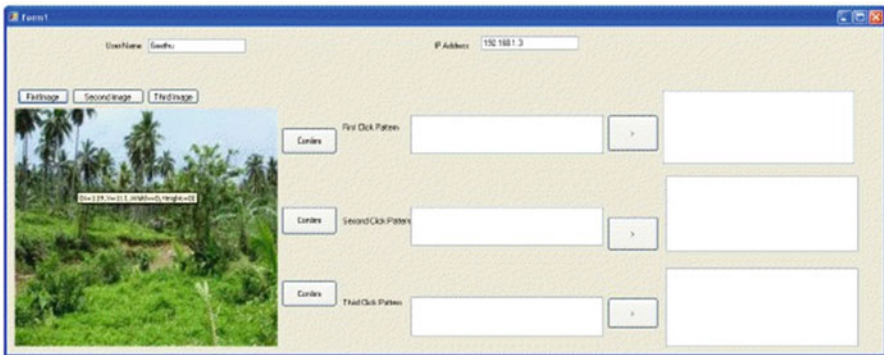


Fig. 42.3 User performing key search operations

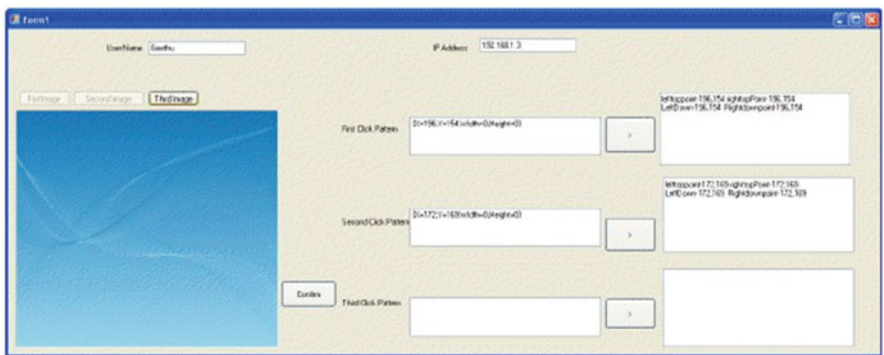


Fig. 42.4 Third-party processing with unique imagery

42.6 Conclusion

The cloud computing trend has generated a lot of concerns worldwide because of its lesser total cost of ownership, aggressive differentiation, scalability, reduced complexity for customers, and quicker and easier acquirement of services. Some believe that cloud is an insecure place for data storage. But the minority people find it safer than their own security provisioning, mainly small businesses that do not have resources to ensure the necessary security themselves. Work mainly included erasure correcting code using file distribution. By utilizing homomorphic tokens with distributed verification of erasure code, the design achieves the integration of data in a secure manner. Our model achieves the desirable properties of securities like dependability, reliability of erasure-coded data, and availability and concurrently identifies the misbehaving servers, its location, and the root cause for misallocating.

References

1. Wang C, Wang Q, Ren K, Lou W (2010) Privacy-preserving public auditing for storage security in cloud computing. In: Proceedings of IEEE INFOCOM'10, March, San Diego, CA, USA
2. Carter L, Wegman M (1979) Universal hash functions. *J Comp Syst Sci* 18(2):143–154. J. Kincaid, <http://www.techcrunch.com>
3. Brately P, Fox BL (2007) Algorithm 659: implementing Sobol's Quasi-random sequence generator. *ACM Trans Math Softw* 14(1):88–100. pp 1–6
4. Erway C, Kupcu A, Papamanthou C, Tamassia R (2009) Dynamic provable data possession. In: Proceedings of CCS, March, pp 213–222
5. Juels A, Burton J, Kaliski S (2003) PORs: proofs of Retrieval for large files. In: Proceedings of CCS '07, pp 584–597, Plank JS, Ding Y (2007) Note: correction to the 1997 tutorial on Reed-Solomon coding
6. Carter L, Wegman M (1979) Universal hash functions. *J Comput Syst Sci* 18(2):143–154
7. Verifying distributed Erasure-coded data. In: Proceedings of the 26th ACM symposium on principles of distributed computing, pp 139–146, 2007
8. Wang C, Chow SSM, Lou W (2011) Privacy Preserving Public Auditing for Secure Cloud Storage. In Ubicsec Lab, IIT, *IEEE Transactions on Computers (TC)*
9. Castro M, Liskov B 2008 Practical byzantine fault tolerance and practical recovery. *ACM Transaction*, June
10. Bowers KD, Juels A, Oprea A (2008) Proofs of retrievability: theory and implementation
11. Wang C, Wang Q, Ren K, Lou W (2009) Ensuring data storage security in cloud computing. In: Proceedings of IWQoS'09, July, pp 1–9
12. Carter L, Wegman M (1979) Universal hash functions. *J Comp Syst Sci* 18(2):143–154. J. Kincaid, <http://www.techcrunch.com>
13. Ateniese G, Pietro RD, Mancini LV, Tsudik G (2008) Scalable and efficient provable data possession. In: Proceedings of the Fourth Int'l Conference in Communication Networks (Secure Comm '08), December, pp 1–10
14. Brately P, Fox BL (1988) Algorithm 659: implementing Sobol's quasi-random sequence generator. *ACM Trans Math Softw* 14(1):88–100
15. Shah MA, Baker M, Mogul JC, Swaminathan R (2007) Auditing to keep online storage services honest. In: Proceedings of 11th USENIX workshop on hot topics in operating systems (HOTOS '07), pp. 1–6

16. Plank JS, Ding Y Note: correction to the 1997 tutorial on reed-Solomon coding. University of Tennessee, tech. Rep. CS-03-504, September 2003
17. Wang C, Ren K, Lou W, Li J (2010) Towards publicly auditable secure cloud data storage services. *IEEE Netw Mag* 24(4):19–24
18. Ren K, Wang C, Wang Q (2012) Security challenges for the public cloud. *IEEE Internet Comput* 16(1):69–73
19. Amazon.com, “Amazon web services (aws),” Online at <http://aws.amazon.com/>, 2009. Smitha Sundareswaran, Anna Squicciarini, “Ensuring distributed accountability for data sharing in the cloud, vol. 9, No. 4, July 2012
20. Wang Q, Wang C, Li J, Ren K, Lou W Enabling public verifiability and data dynamics for storage security in cloud computing, Sep 2009.
21. Wang Q, Ren K, Lou W, Zhang Y Dependable and secure sensor data storage with dynamic integrity assurance. In: *Proceedings of IEEE INFOCOM’09*, Rio de Janeiro, Brazil, April 2009.
22. Hendricks J, Ganger G, Reiter M (2007) Verifying distributed erasure-coded data. In *26th ACM*, March

Chapter 43

Agile Supply Chain Management Enabled by the Internet of Things and Microservices



G. Selvakumar and L. S. Jayashree

Abstract The business networks of modern world span across the globe, and their requirements for supply chain management and logistics have grown extensively. The field is continuously innovating to make the supply chain more predictable, to optimize the logistics, and to be cost-efficient. This paper investigates the impact of using the Internet of Things in supply chain management in the first part. The software engineering methodologies are rapidly evolving to meet the current business requirements of the modern world, and companies are moving toward more agile methodologies. Microservices architecture solves numerous problems in the modern software development process. The impact of following microservices architecture is analyzed in the second part of the paper. Finally, all the dots are connected by discussing the development of IoT-enabled supply chain solutions with microservices architecture. While microservices, IoT, and supply chain management are evolving in their own aspects, this paper is an attempt to bring all the three together to achieve more agile supply chains to satisfy the growing business needs.

Keywords Supply chain · Microservices · DevOps · Internet of Things · Agile · Monolithic architecture

Abbreviations

SCM Supply chain management
IoT Internet of Things
REST REpresentational state transfer

G. Selvakumar (✉)

Department of Computer Science, Sri Shakthi Institute of Engineering and Technology,
Coimbatore, Tamil Nadu, India

L. S. Jayashree

Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_43

API Application programming interface
UX User eXperience

43.1 Introduction

Every industrial sector has started analyzing the ways of utilizing the advancements in cyber physical solutions in their operations. IoT promises huge results in terms of revenue, safety, monitoring, and management in every business unit of an enterprise.

The Internet of Things is found to be useful for three primary reasons [1].

Cost Reduction The end-to-end process management can be administered very well in all the stages. General administration, supply chain management, design and production departments, distribution, sales, and all other functional units can be administered effectively.

Quality Improvement The cost of quality can be minimized by employing more intelligent predictive maintenance techniques.

Regulatory Compliance Availability of data from every corner makes an information-rich environment which reduces the cost of conformance.

Among several sectors, supply chain management is the one in which plenty of advancements happen very frequently due to its impact on the nation's economy and participating companies' bottom line.

The purpose of this chapter is to combine the advancements in software engineering practices and Internet of Things technologies to meet the challenges of today's supply chain.

43.2 Background

The Internet of Things plays a pivotal role in running several businesses smoothly. The one area in which the impact of information services is very predominant is supply chain and logistics. The technology plays a major differentiator among the companies, and every supply chain and logistics provider is keen on ensuring that they make use of the latest technology stack for information processing. Along with the goods, information flows across various ends of the supply chain, and the Internet of Things is the major actor that ensures that appropriate information is collected and distributed throughout the lifecycle of supply chains [2]. Managing the warehouses or managing the shipments and logistics or managing the people across supply chains, the Internet of Things has made the impact so strongly in the recent past.

The Internet of Things promises to manage every single detail like monitoring energy bills, keeping the wastage under control, tracking the consignments along with their quality, and a lot more similar things. The collected data is mostly sent to the cloud for analysis and to produce recommendations which would support critical decision-making.

In this whole process, we can clearly observe that the software development strategy for supply chain companies is quite different from other domains. The data collection, analysis, and dissemination of results must happen as fast as possible, since the associated goods and human workforce are always on the move, and a small delay may cause a huge wastage or loss.

So, like every industry, the innovative software development practices are being tried in supply chain management also. This chapter is an attempt to focus on the unique challenges of supply chain management and logistics domain and the ways to apply modern software development processes to face the challenges.

43.3 Empowering Supply Chains with IoT

Every global industry has been affected by the Internet of Things technology in almost all of their major areas of operations. Beyond the manufacturing sector, automobiles, healthcare, and retail are some of the major industries which experienced the impact of IoT in the beginning. The most exciting part is, in all those industrial sectors, supply chain management plays a crucial role in their day-to-day operations. Several major players are striving toward new revenue opportunities and maximum operational efficiency through Internet of Things applications. Initially it was considered as a way to differentiate from the competitors. However, the supply chain companies have understood shortly that the Internet of Things could completely change their business model.

The operational efficiency of an enterprise largely depends upon the relationship among various trading partners such as suppliers, vendors, or customers. The IoT improves the quality of communication among the various stack holders which ultimately enhances the quality of business and the relationships. The manufacturing industry makes use of the Internet of Things tremendously in their supply chains to track their assets. The tracking numbers and bar codes are no longer very effective. The products from the manufacturing floor to the retail store can be tracked using GPS sensors and RFID tags. Not only tracking the products, significant data such as temperature, pressure, humidity, and variation in similar parameters on its transportation also can be obtained from the sensors [3]. The supply chain and logistics companies always strive for on-time deliveries, better quality control mechanism, and forecasting the product demand. The data gained from the IoT ecosystem helps a lot in this regard.

The inventory control systems have become extremely powerful with the help of IoT. If the manufacturing company could track the inventory including the supplies

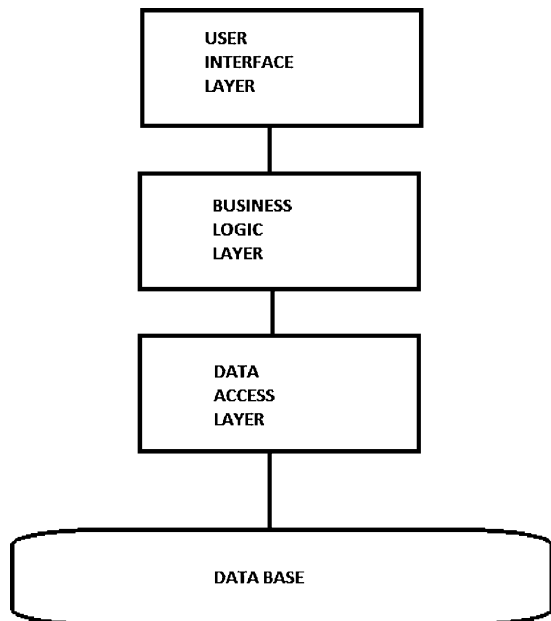
for future manufacturing very conveniently on their mobile screens, there would no longer be cases of missing deadlines. All the data gained through the IoT can even be used to analyze the trends and optimize the schedules in manufacturing and logistics.

43.4 Agile Development with Microservices

The monolithic architecture is the conventional way we build software systems in which the outcome is a single software application and the various functionalities are interwoven with each other. Every module will have dependency to other module in terms of function calls or data management [4]. They would be following mostly the same technology stack for interoperability purpose and the same standards across the application. So, almost all the initial solutions for enterprise IoT were designed (Fig. 43.1).

It is essential to understand the defects of monolithic architecture before we proceed further. Assume that we have an enterprise IoT solution developed as a monolithic one and we want to improve the system performance or we want to scale it according to the growing number of users. This is possible only by multiple deployments. Moreover the specific functionalities which are overloaded as per the changing business requirements will create a bottleneck [5]. This will cost a huge waste of resources. Saying simply, the whole application is united as a single entity,

Fig. 43.1 Monolithic architecture



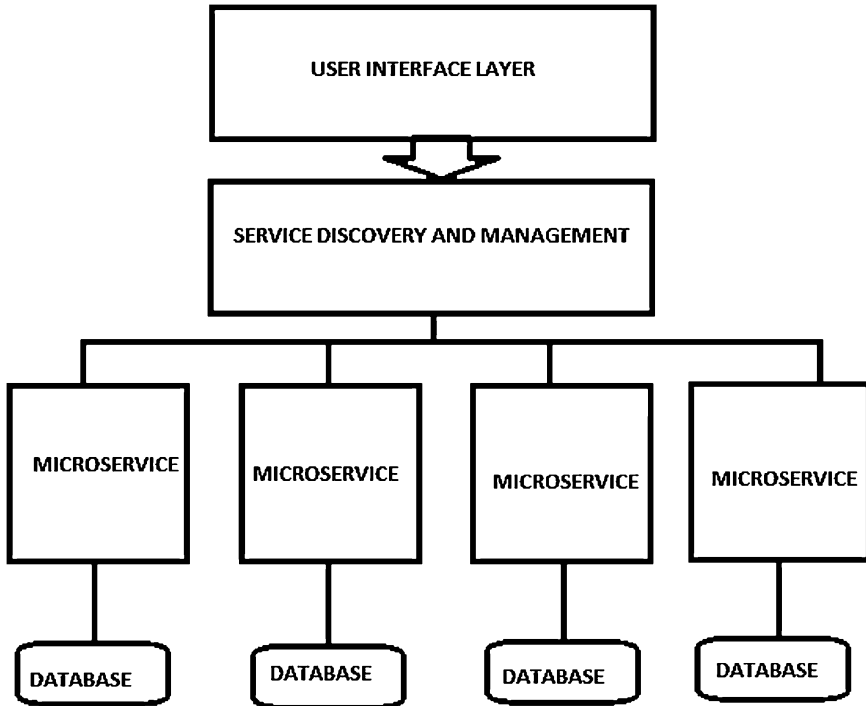


Fig. 43.2 Microservices architecture

and we cannot modify or enhance specific components of the application independently. Maintenance of the application and redeployments will become harder and harder as the system grows, and it will make continuous integration a more challenging one. Finally, as a single technical entity, there is very less scope to experiment with new technical stack, and the solution should be developed only in a specific technology stack. Thus we have only a limited number of options to solve a business problem, and we cannot experiment with completely independent technology platforms.

Instead of in-memory function calls, separate processes will be implemented. These are functionally atomic and highly decoupled and task specific. They communicate with each other using lightweight mechanisms like REST API (Fig. 43.2).

Loose coupling among the components of a solution is always considered to be the most important software engineering metric which ensured flexibility. It makes the system a light-weighted one with maximum maintainability [6]. Microservices ensure loose coupling, and it makes it possible to integrate with new technologies such as Big Data, Geo service, or automation.

43.5 Microservices for IoT Applications

More and more enterprises started to adopt enterprise IoT, and the applications are getting much larger gradually. It increases their complexity also which results in difficulties in scaling the systems and maintaining them. Though there are multiple business units, stakeholders, and functional components, traditionally we build systems in a monolithic way [5].

In the field of IoT applications, microservices architecture has been introduced in response to the challenges faced by monolithic applications. So, almost all the initial solutions for enterprise IoT were designed in terms of monolithic architecture principles.

While we discussed the nature of microservices compared to monolithic architecture in the previous section, we investigate the need for using microservices in IoT applications in this section. The ecosystem of IoT applications especially in enterprises are so complex due to the involvement of a variety of sensors, data format, servers, and different applications. Their interfaces to the real world also is multi-faceted like web, mobile, tablets, dedicated kiosks, etc. Quality assurance systems in each layer also have to be developed according to the nature of the layer, eventually ensuring the overall performance of the IoT application. Even conventional agile practice needs huge efforts in integrating all these devices, applications, servers, and data. Already the companies spent a huge proportion of their budget for IoT systems only for integration. With the help of microservices, these complicated applications can be broken down to independently deployable services. The integration complexity is reduced a lot, and a huge application can be built and maintained.

The requirements from various business users in terms of UX design and precise functionality and performance metrics demand frequent changes to several components of the platform. Using DevOps and Microservices together ensures that the change requests can be handled with ease due to continuous deployment and integration. Any modification in the device level or user interface level or in the business logic can be done quickly without huge impact on other parts of the application.

43.6 Connecting the Dots: IoT-Enabled Supply Chain Solutions with Microservices

Now we can analyze the significance of connecting the dots discussed above. The combination of IoT and microservices is required for supply chain solutions due to the specific challenges mentioned below.

There is always a lack of coordination between the operations and IT team due to their inherent nature of working. While the operations team deals with the day-to-day challenges of business, the IT team may not be able to closely follow the changes

in the operations due to the development life cycle conventionally followed in the IT. Microservices and DevOps are the trends slowly changing the situation and making application development more agile. It offers tremendous operational efficiency for the enterprises. Microservices-based development cycle supports continuous integration which results in frequent software releases [7]. This enables continuous delivery which speeds up the deployments in response to the changes in the requirements. While DevOps talks about common practices to be followed for more agile development, microservices defines the toolset and architectural best practices. Combined together, they work great in building more agile solutions. As discussed in the previous section, developing solutions for supply chain management especially with IoT enablement can get very complicated due to the involvement of several companies and time-critical IT systems. This results in multiple stages of supply chains with integration challenges and other vulnerabilities [8]. Thus microservices architecture is the natural choice for IoT-enabled supply chain management.

As previously illustrated, multiple stakeholders and systems are participating in the supply chains, and the interoperability has always been an issue. The same kind of skill expertise cannot be expected at various stages. The learning curve required to make the interoperability will be huge, and it will not be suitable for the business requirements which are continuously changing and technology upgrading. Microservices yields a great deal of freedom in choosing any technology stack for a given business service. Thus the teams across the supply chain would work independently on their part which all would be integrated seamlessly [9].

43.7 Conclusion and Future Work

This paper is an attempt to view the impact of developing scalable and flexible applications with the help of microservices, especially in the domain of supply chain management. The significance of using the Internet of Things in supply chain management was discussed, and the challenges in developing software applications in this sector were analyzed. The analysis makes it very clear that the development of IoT-enabled supply chains is complicated, and we need some innovative software development practices to meet the challenges. Then, a clear analysis of modern software development strategy with microservices was given, and it was compared to the conventional monolithic architecture. In the final part, it has been clearly illustrated how microservices can be used to make IoT-enabled supply chain systems more agile. The future scope of this work is to develop a prototype of a microservices-based application for one particular type of supply chain, say food supply chain, and to prove the impact of using microservices to make the whole development more agile.

References

1. Krylovskiy A, Jahn M, Patti E (2015) Designing a Smart City internet of things platform with microservice architecture. In: 3rd international conference on future internet of things and cloud, Rome, Rome, pp 25–30
2. Cortés B, Boza A, Pérez D, Cuenca L (2015) Internet of things applications on supply chain management in: world academy of science, engineering and technology. *Int J Comp Info Eng* 9 (12):889–895
3. Lianguang M (2014) Study on supply-chain of agricultural products based on IOT. In: sixth international conference on measuring technology and mechatronics automation, Zhangjiajie, pp 627–631
4. Tu M (2018) An exploratory study of Internet of Things (IoT) adoption intention in logistics and supply chain management – a mixed research approach. *Int J Logist Manag* 29(1):131–151
5. Shadija D, Rezai M, Hill R (2017): Microservices: granularity vs. performance. In: Proceeding UCC '17 companion proceedings of the 10th international conference on utility and cloud computing, pp 215–220
6. Innerbichler J, Gonul S, Damjanovic-Behrendt V, Mandler B, Strohmeier F (2017): NIMBLE collaborative platform: microservice architectural approach to federated IoT. *Global Internet of Things Summit (GloTS)*, Geneva, pp 1–6
7. Lou P, Liu Q, Zhou Z, Wang H (2011) Agile supply chain management over the internet of things: In: international conference on management and service science, Wuhan, pp 1–4
8. Kant K, Pal A (2017) Internet of perishable logistics. *IEEE Internet Comput* 21(1):22–31
9. Riad M, Elgammal A, Elzanfaly D (2018) Efficient Management of Perishable Inventory by utilizing IoT. In: 2018 IEEE international conference on engineering, technology and innovation (ICE/ITMC), IEEE, Stuttgart, pp 1–9

Chapter 44

Production and Characterization of Bio-Fertilizers from Tree Leaves Utilizing an Automated Hot Composting Chamber with Cyber-Physical Systems



**Mahendran Rajagopalan, Vijayakumar Arumugam, Uma Dharmaligam,
Kavinilavu Anbalagan, and Anupriya Chandrasekaran**

Abstract Composting is a reliable process to transform waste dry leaves into superior quality compost. In this research, we proposed a rapid composting technique for converting powdered dry leaves to mature compost in an automated rotary drum composter and for monitoring the physicochemical properties of the pile by using cyber-physical systems. Here, the solar energy-based equipments were utilized for grinding dry leaves and energizing a heating coil inside the composter. The optimal temperature of the composter had been optimized and maintained at 50–55 °C by using closed-loop heating coils and monitored by using temperature sensors. During composting, the moisture content (MC) decreased with respect to the duration. On the contrary, the pH and electrical conductivity (EC) gradually increased as the dried leaves are being converted to mature compost. The aim of this research is to automate the composter by using sensing devices and monitor the physicochemical characteristics in terms of pH, electrical conductivity, moisture content, and temperature. Here, we anticipated a cost-effective, less maintenance, and eco-friendly quick-composting technique for obtaining good quality of compost from the leaves within 30 days.

Keywords Rotary drum composting · Physicochemical parameters · Dry leaves · Process monitoring · Compost quality · Cyber-physical systems

M. Rajagopalan · V. Arumugam (✉) · K. Anbalagan · A. Chandrasekaran
Anjalai Ammal Mahalingam Engineering College (Anna University), Kovilvenni, Tamil Nadu,
India
e-mail: hodeee@aamec.edu.in

U. Dharmaligam
School of EEE, SASTRA Deemed University, Tirumalaisamudram, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_44

Abbreviations

N-P-K	Nitrogen-phosphorous-potassium
MC	Moisture content
EC	Electrical conductivity
w/v	Weight/volume

44.1 Introduction

With respect to the crop production level in India, it is projected that the crops produce around 679 million tonnes of residues every year, and the anticipated N-P-K supplies from the crop residues are 6.25 and 10.25 million tonnes for the years 2011 and 2030, respectively [1, 2]. Unfortunately, around 90% of these wastes are used for land filling and uncontrolled dumping on the outskirts of cities, which cause serious environmental problems in global warming by the emission of greenhouse gases [2]. Hence, an effective technology is required to compost the crops' residue into a prolific product in a short period of time. Composting is one of the effective techniques to convert crop residue to good nutritional quality compost [3, 4]. It is an exciting global process to turn wastes into a useful product; however it depends on the cost, nature of the pile material, and the amendment of the produced compost to the soil.

Composting is the biological decomposition and stabilization of the organic materials into a dark rich substance, which is stable, nonpolluting, and free of pathogens and possesses a significant value in agriculture [5]. It is a vital process in waste management, owing to its robustness and obtaining valuable products with soil amendment potential. Furthermore, it is beneficial to the land and ecosystems, which include as soil conditioners, as fertilizers, addition of humic acids to the soil, as natural pesticides, in erosion control, in land reclamation, in wetland construction, and as landfill cover [4, 5]. Specifically, the quality of the compost depends upon the source, composting facility design, composting technique, and duration of the composting process. Generally, the composting process occurs in two phases. Firstly, the decomposition phase (breaking of complex organic matter) has three stages, namely, mesophile, thermophile, and cooling. Secondly, the humification phase, which is related to the maturation phase, is characterized by the reorganization of the organic matter in the pile [6, 7]. In particular, the compost from the dried leaves has certain advantages such as economically feasible, eco-friendly, and publicly acceptable [8]. Generally, the efficiency of the composting depends upon the pH, carbon-to-nitrogen ratio, moisture content (MC), ionic exchange capacity, aeration in the chamber, bulk density, particle size, and porosity of the composting piles [9–11]. For rapid composting, the leaves must undergo size reduction in order to increase the surface area, which enhances the speed of biological oxidation with

respect to the quantity of surface [6]. Furthermore, the maturation is carried out at ambient temperature through the mesophilic microorganisms (bacteria and fungi) [6]. The conversion of mature compost can be detected by a pleasant odor, growth of white fungi, low temperature, absence of ammonia odor, granular texture, reduction of pile volume, and color change of the pile into brown [12, 13].

Conventionally, composting is carried out by using waste organic materials and waiting for the materials to split into humus after several months [14]. Conventional composting has certain drawbacks such as utilization of space for a long time, leaching out of some nutrients during rainfall, and production of disease-causing organisms and insects [15]. In our research, we overcame the above drawbacks by utilizing a fully automated rotary drum composter, which possess significant benefits, for instance, inexpensive system, better moisture-holding capacity, superior air circulation, and effortless stirring of the composting materials by rotating the chamber.

Rotary drum composting is traditionally carried out by cold process, which possess definite drawbacks such as uneven decomposition, nutrient leaching, pest attraction, anaerobic decomposition, and very slow process (10–12 months) [16]. In order to overcome the above drawbacks, we are going to employ hot composting process, which produces the mature compost in a short period of time. Furthermore, it has significant benefits such as killing weed seeds and disease-causing pathogens, breaking down the organic materials into fine humus, need less space, no pests, degradation of toxin, and more efficient process. However, hot composting acquire certain difficulties, for example, it needs a lot of manual effort, fire hazard, and death of microbes at a high temperature [17]. In order to prevail over the mentioned drawbacks, in this research, we proposed a fully automated system with heating coils for optimizing and controlling the temperature inside the chamber and monitored the physicochemical characteristics (temperature, moisture, pH, and ionic conductivities) by specific sensing devices, which are essential parameters for effective composting of the pile. In addition, the temperature and humidity of the atmosphere were also measured using sensors.

44.2 Materials and Methods

44.2.1 Sampling

Dry tree leaves (*Millettia pinnata*) were collected from our (AAME) college campus at Kovilvanni, Tiruvarur, India. The leaves were cleaned in running water and dried in sunlight for a week. Subsequently, the dried leaves were chopped by using a grinder followed by sieving using a mesh. The grinding of leaves was carried out by using a specially designed grinding system (Fig. 44.1).

Fig. 44.1 Image of the grinding system



Fig. 44.2 Image of the composting chamber



44.2.2 *Design of Composting Chamber*

The composting process was carried out in a plastic drum of 125 L capacity in batch mode process, and the drum was suitably customized for air circulation. Additionally, black color paint was applied on the outer surface of the rotary drum for retaining the heat from the sunlight. A sampling window at the middle part of the chamber was used to feed the powdered dry leaves (~15 kg) and used to collect the samples for physicochemical analysis, periodically. For draining out the leachate, a hole was made under the drum. During composting, rotation of the composting chamber had been carried out from time to time. The image of the rotary composting chamber is shown in Fig. 44.2.

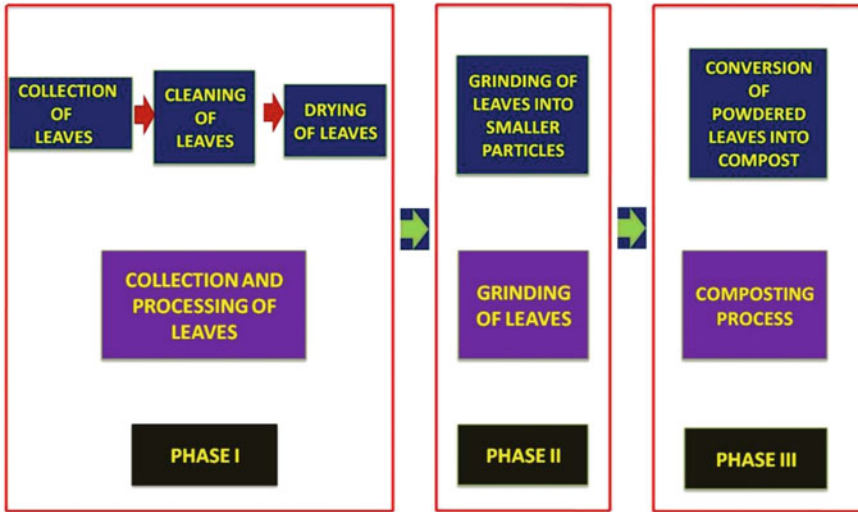


Fig. 44.3 Schematic of the proposed composting process

44.2.3 Physicochemical Parameters of Composting Materials

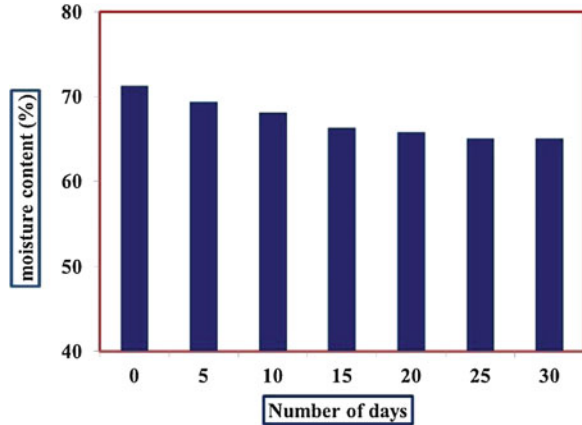
Physicochemical properties such as temperature, moisture, electrical conductivity (EC), and pH (1,10 w/v) were measured during the composting process by using specific sensors such as digital thermometer and humidity, electrical conductivity, and pH meters, respectively. The samples were manually collected from three spots in the chamber and mixed thoroughly to make a homogenized sample for analysis. Additionally, the moisture content of the powdered dry leaves to the mature compost was found out by using the gravimetric method [18]. Furthermore, solar radiation, temperature, and humidity of the atmosphere had been measured with specific sensors every day. Specifically, we employed a photovoltaic system, comprising solar cells used to generate electrical power for grinding the leaves and heating the coils inside the chamber. The schematic representation of the proposed composting process is shown in Fig. 44.3.

44.3 Results and Discussion

44.3.1 Moisture Content

Moisture content (MC) is a critical factor which influences the oxygen uptake rate, free air space, microbial activity, and temperature and the optimum MC for effective composting mainly depends on the type and nature of the composting material. For

Fig. 44.4 Changes in the moisture content during the composting period

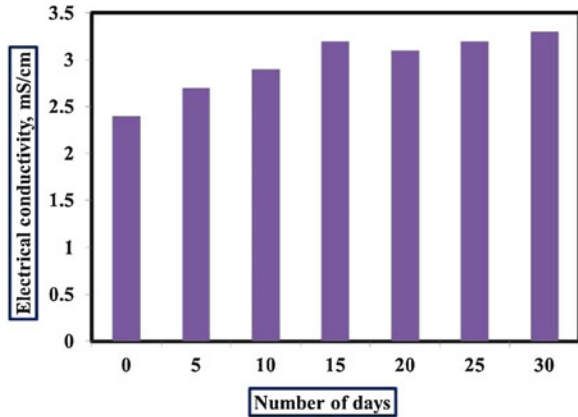


optimizing and enhancing the composting process, the decomposition of the organic materials by the microbial activities is a significant one and the composting works well if the moisture content of the pile is around 50% [18, 19]. Specifically, very low MC causes early dehydration and seizes the composting, which produces a biologically unstable product [19]. On the contrary, high MC produces soggy mass, clumpy, and anaerobic conditions as a result of water logging and high transportation cost, which prevent the composting process [19]. In this research, the MC gradually decreased from 71.3% to 65.1%, from the powdered dried leaves to the matured compost, respectively (Fig. 44.4), owing to maintain the optimal temperature of 50–55 °C during the composting process.

44.3.2 Influence of Temperature

Generally, the temperature of the chamber maintain around 50–55 °C for an effective composting process. The optimal temperatures are used to achieve proper sanitation, rapid degradation, water evaporation, and humification during composting [11, 19]. In particular, high temperature should be avoided because they slow down the biological activity and cause undesirable chemical modifications of the composting pile. Additionally, heat is also generated by the respiration of the microorganisms by breakdown of the organic materials [6]. In this research, the compost chamber had been frequently rotated to prevent the pile from getting too hot [20]. Turning the pile is not only used to reduce overheat but also used for well aeration, which is essential to decompose organic materials in the chamber. During composting, the thermophilic phase (> 45 °C) occurs for a longer time in the composting process as it attributes toward the killing of pathogens and reduces biodegradable carbon [6]. Furthermore, the development of eumycetes and actinomycetes, which are the main decomposers of long-chain polymers, cellulose, and lignin, functioned well at the optimum temperature (50–55 °C) [6]. In this research,

Fig. 44.5 Variation of electrical conductivity during composting



we optimize, maintain, and monitor the temperature by using a digital temperature sensor, maintained around 50–55 °C throughout the process by using the closed-loop heating system.

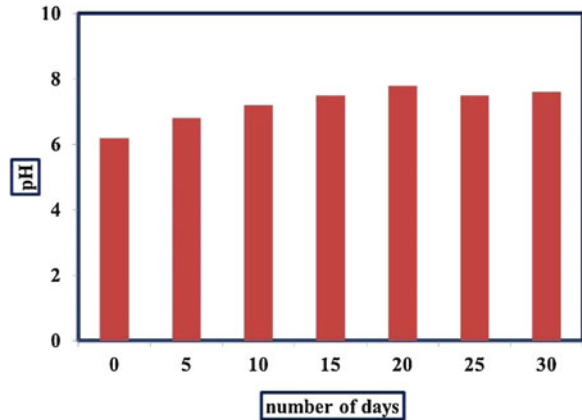
44.3.3 *Electrical Conductivity*

The electrical conductivity (EC) of the dried leaves was 2.4 and the mature compost was 3.3 S/m (Fig. 44.5); the EC measurement is used to identify the presence of soluble salts during composting. The enhancement of the EC was observed during thermophilic phase of the composting material owing to the decomposition of organic matter into the nutrients [6]. Significantly, augmentation of the EC during composting was mainly due to the increase of concentration of cations in the chamber [7]. In particular, the humic compounds formed during composting possess a high capacity of swap over positively charged ions, which are easily exchanged with other cations. The EC tends to increase during composting, as the substrates are humified, and forms carboxyl and phenolic functional groups [21]. The EC values reflect the degree of salinity and illustrated its possible phytotoxic effects on plant growth when applied to the soil.

44.3.4 *pH Measurements*

pH is a vital factor which influences the degradation process of the organic matter and affects the activities of microbes during composting [10]. Generally, optimal pH values are in between 5.5 and 8 for an effective composting process, and the pH values around neutrality are most favorable for the development of microorganisms in the pile [6]. Furthermore, the pH may be considered as good indicator for compost

Fig. 44.6 pH variation during the composting process



stability and maturity [14]. In this research, the pH varied from 6.2 to 7.6 for the dried leaves and mature compost, respectively. The pH of all the trials at regular time intervals was in the range of 6.2–7.6. It might be due to the dried leaves possessing a high nitrogen content, which favors the conversion of nitrogen-bound organic compounds into ammonia at the final stage of the composting process [6, 21, 22]. Here, the improvement of pH was due to the activity of proteolytic bacteria releasing ammonia, by breakdown of proteins and nitrogenous bases. Further, the pH enhancement depends on the conditions of airflow in the chamber, which allows degradation of the dried leaves at higher pH [10] (Fig. 44.6).

44.4 Conclusions

The obtained compost had a dark brown color and an earthy odor. The moisture content decreased during the composting process owing to maintain optimal temperature range during the composting process. On the contrary, the pH and EC slightly increased from the dried leaves to the mature compost. Specifically, the grinding of the leaves into fine particles and maintaining the temperature by using heating coils inside the composting chamber were powered by solar. The optimal moisture content was 65.1%, pH was 7.6, and electrical conductivity was 3.3 mS/cm for the mature compost after 29 days. Hence, the proposed rotary drum composter is an effective technique for rapid composting of dried tree leaves.

Acknowledgments The authors would like to thank the Tamil Nadu State Council of Science and Technology for the financial support for this research.

References

1. Lin L, Xu F, Ge X, Li Y (2018) Improving the sustainability of organic waste management practices in the food-energy-water nexus: a comparative review of anaerobic digestion and composting. *Renew Sust Energy Rev* 89:151–167
2. Esmailian B, Lewis B, Duarte F, Rattif C, Behdad S (2018) The future of waste management in smart and sustainable cities. *Waste Manag* 81:177–195
3. Proietti P, Calisti R, Gigliotti G, Nasini L, Regni L, Marchini A (2016) Composting optimization: integrating cost analysis with the physical-chemical properties of materials to be composted. *J Clean Prod* 137:1086–1099
4. Papalia T, Settineri G, Mallamaci C, Jeske- Kaczanowska A, Muscolo A (2018) Are raw materials or composting conditions and time that most influence the maturity and/or quality of composts? Comparison of obtained composts on soil properties. *J Clean Prod* 195:93–101
5. McMahon V, Garg A, Aldred D, Hobbs G, Smith R, Tothill I (2009) Evaluation of the potential of applying composting/bioremediation techniques to wastes generated within the construction industry. *Waste Manag* 29:186–196
6. De Bertoldi M, Vallinia G, Pera A (1983) The biology of composting: a review. *Waste Manag Res* 1:157–176
7. Azim K, Soudi B, Boukhari S, Perissol C, Roussos S, Alami T (2018) Composting parameters and compost quality: a literature review. *Org Agric* 8:141–158
8. Kalamdhad AS, Singh YK, Ali K, Kazmi M (2009) A rotary drum composting of vegetable waste and tree leaves. *Bioresour Technol* 100:42–50
9. Guo R, Li G, Jiang T, Schuchardt F, Chen T, Zhao Y, Shen Y (2012) Effect of aeration rate, C/N ratio and moisture content on the stability and maturity of compost. *Bioresour Technol* 112:171–178
10. Iglesias E, Victor J, Garcia P (1991) Composting of domestic refuse and sewage sludge. I. Evolution of temperature, pH, C/N ratio and cation-exchange capacity. *Resour Conserv Recycl* 6:45–60
11. Suler DJ, Finstein MS (1977) Effect of temperature, aeration, and moisture on CO₂ formation in bench-scale continuously thermophilic composting of solid waste. *Appl Environ Microbiol* 33:345–350
12. Tiquia SM (2005) Microbiological parameters as indicators of compost maturity. *J Appl Microbiol* 99:816–828
13. Namkoong WE, Hwang Y, Cheong JG, Choi JY (1999) A comparative evaluation of maturity parameters for food waste composting. *Compost Sci Util* 7:55–62
14. Wichuk K, McCartney D (2013) Compost stability and maturity evaluation—a literature review. *J Environ Eng Sci* 8:601–620
15. Wu L, Ma L, Martinez G (2000) Comparison of methods for evaluating stability and maturity of biosolids compost. *J Environ Qual* 29:424–432
16. Jain M, Kalamdhad AS (2018) Efficacy of batch mode rotary drum composter for management of aquatic weed (*Hydrillaverticillata*). *J Environ Manag* 221:20–27
17. Singh YK, Kalamdhad A, Ali M, Kazmi AA (2009) Maturation of primary stabilized compost from rotary drum composter. *Resour Conserv Recycl* 53:386–392
18. Guo R, Li G, Jiang T, Schuchardt T, Chen T, Zhao Y, Hen YS (2012) Effect of aeration rate, C/N ratio and moisture content on the stability and maturity of compost. *Bioresour Technol* 112:171–178
19. Agnew J, Leonard J (2003) The physical properties of compost. *Compost Sci Util* 11:238–264
20. Tiquia S, Tam N, Hodgkiss I (1997) Effects of turning frequency on composting of spent pig-manure sawdust litter. *Bioresour Technol* 62:37–42
21. Goyal S, Dhull S, Kapoor K (2005) Chemical and biological changes during composting of different organic wastes and assessment of compost maturity. *Bioresour Technol* 96:1584–1591
22. Sundberg C, Yu D, Whittle I, Kauppi S, Smars S, Insam H, Romantschuk M (2013) Effects of pH and microbial composition on odour in food waste composting. *Waste Manag* 33:204–211

Chapter 45

Spectrum Sensing Based on Cascaded Approach for Cognitive Radios



N. Iswarya and L. S. Jayashree

Abstract The swift growth in radio communication technology has drawn to scarcity in wireless spectrum. Literature points out that licensed spectrums allotted by regulatory agencies are underutilized in the allocated band of spectrum. Cognitive radio networks appear to be a promising solution to address the bandwidth scarcity and demands of wireless spectrum. Cognitive radios are confronted to maximally utilize the spectrum through sharing of spectrum with the licensed primary users (PU). The efficiency of cognitive radios mainly depends on the efficiency of the spectrum sensing plane, in which better spectrum utilizations are exploited. A hybrid approach of merging energy detection (ED)-based channel sensing and cyclic prefix autocorrelation detection (CPAD) techniques has cascaded in the way to boost the probability of detection, which is proposed in this chapter. Energy detection techniques implicate no prior knowledge of PU signals, less computational complexity, and low energy consumption but hold uncertainties at low SNRs (at -20 dB to 10 dB). The next mentioned CPAD technique is more robust at less decibels, but it requires a large number of samples yielding complexity and increase in sensing time. Consequently, ED and CPAD techniques, with an influence on the benefits of each technique, have been designed as cascading and implemented using the Universal Software Radio Peripheral (USRP) tool. On comparison, the probability of the detection of ED and CPAD at SNR ranging from -20 dB to 5 dB is around 0.3 – 0.9 and 0.6 – 0.9 , respectively. The cascaded design has the probability of detection in the bound 0.7 – 1 for the same specified SNR. Spectrum sensing based on cascading a couple of detectors outperforms in detection probability compared to a single detector.

Keywords Cognitive radio · Spectrum sensing · Energy detection · Cyclic prefix autocorrelation detection

N. Iswarya (✉) · L. S. Jayashree
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

Abbreviations

AWGN	Additive white Gaussian noise
CD	Covariance-based detection
CPAD	Cyclic prefix autocorrelation detection
CSD	Cyclo stationary feature detection
ED	Energy detection
EME	Energy with minimum eigenvalue ratio
MD	Matched filter detection
MME	Maximum to minimum eigenvalue ratio
PU	Primary users
RF	Radio frequency
SCD	Spectral correlation density
SDR	Software-defined radio
USRP	Universal software radio peripheral

45.1 Introduction

Studies have shown that the radio spectrum is underutilized by the users continually to whom the resources are allocated to. Utilization of the spectrum is evaluated to be in the vicinity of 5% and 85% on an entire range of licensed radio frequency (RF) bands at different geographical areas and at distinct periods. Significantly cognitive radio, introduced by Mitola in 2000, has turned up to be a promising solution to improve spectrum utilization. A cognitive radio, built on a software-defined radio (SDR), is an intelligent heuristic wireless communication system to establish highly reliable communications as required and to ensure the proficient utilization of the radio band. The technology determines accessible portions of licensed spectrum and utilizes such portions for secondary use. Through a spectrum sensing task, the radio frequency spectrum is observed, and these perceptions are processed to determine the channel occupation by an authorized user [1]. This task should examine the appearance or nonappearance of the primary user (PU) signal reliably within a restricted sensing time. Regardless of estimating the spectral content or the radio frequency vitality over the spectrum range, at the point when cognitive radios are subjected, spectrum sensing is a broader term that includes acquiring the spectrum qualities over multiple dimensions such as space, time, frequency, and code. It additionally includes figuring out the types of signals that occupy the spectrum along with modulation, waveform, bandwidth, carrier frequency, etc. In any case, this needs dynamic signal investigation procedures with extra-computational multi-faceted nature. Some of the spectrum sensing techniques proposed during the recent decade include energy detection (ED) [2], matched filter detection (MD) [3], cyclostationary feature detection (CSD) [4], and covariance-based detection (CD) [5] (Fig. 45.1).

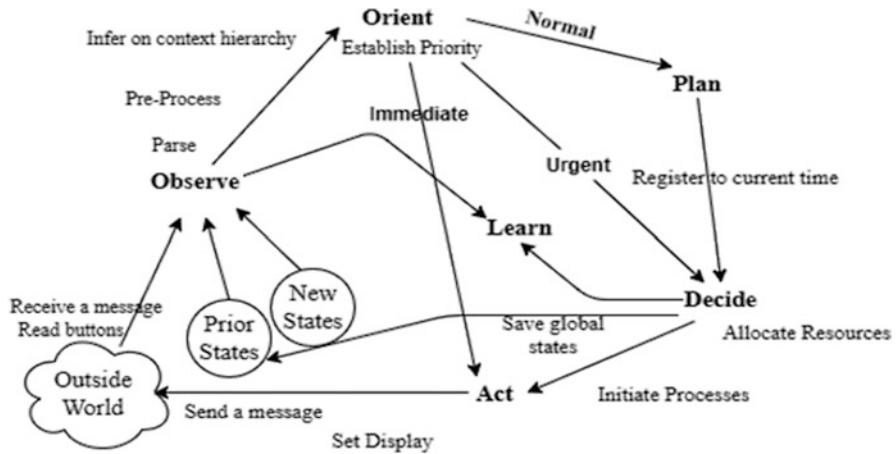


Fig. 45.1 A cognitive cycle. (Image from [5])

45.2 Related Work

The energy detection technique is the simplest spectrum sensing approach used in cognitive radio [2]. Although it requires no prior knowledge about the primary signal, it does require knowledge of the noise power. In this method, the decision on the occupancy of the spectrum is made by comparing the average energy of the observed samples with a predetermined threshold that depends on the noise floor. In matched filtering, a matched filter correlates the known primary user signal with the received signal to detect the presence of PU. The matched filtering needs few samples; therefore, it requires short time to achieve low probabilities of missed detection and false alarm [3]. Cyclostationary detection takes advantage of the fact that statistical parameters of practical communication signals vary periodically [4]. Cyclostationary data can be extracted from the spectral-correlation density (SCD) function. Thus, it is clear that both MF and CSD need information about the primary user signal prior to receiving it. Covariance spectrum sensing can overcome these problems, since it capitalizes on the fact that the co-variance matrix of only noise samples and the one of signal plus noise samples behave differently; therefore, it requires no information about either the noise or the signal power [6]. The co-variance matrix of the samples collected by the receiver contains information exploitable for spectrum sensing. The authors introduce two statistics: the sum of the matrix elements that are not in the main diagonal, and the sum of the elements that are in the main diagonal. The comparison of the ratio between these two statistics with a threshold can indicate the presence of either signal plus noise or only noise. In [7] the authors propose other two metrics: the maximum-to-minimum

eigenvalue (MME) ratio and the average received power to minimum eigenvalue ratio, also referred to as the energy with the minimum eigenvalue (EME) ratio. The paper is structured as follows: A cascaded algorithm of energy and autocorrelation-based detection is described in Sect. 45.3. Section 45.4 shows the simulation results of the proposed methods. Finally, Sect. 45.5 concludes the paper.

45.3 Detection Method

An ideal sensing detector has to know the estimations of primary user signal, noise, and channel gain. In practice, there is no information about some or all of these parameters. Among different detection methods, the energy detector and the autocorrelation-based detector do not require prior information of the primary user signal, but its computation time can be reduced if it has knowledge of symbol time T_d .

45.3.1 Energy Detection

Energy detection can be applied for any signal type since it does not require prior information of the primary user. A conventional energy detector consists of a low-pass filter to reject out of band noise and adjacent signals, Nyquist sampling A/D converter, square law device, and integrator. A complex base band equivalent to the energy detector is considered. H_0 and H_1 are the hypotheses of the absence and presence of the primary user signal, respectively. The energy detector makes decision based on M observations. The signal x_k is the input signal, where $k = 1, 2, 3, \dots, M$

$$x_k = \begin{cases} n_k, & \text{under } H_0 \\ s_K + n_k, & \text{under } H_1 \end{cases} \quad (45.1)$$

with the primary user signal as s_K and noise as n_k . The noise n_k and the primary user signal s_K are assumed to be independent and identically distributed random Gaussian processes with zero mean. The variance for the primary signal is σ_s^2 and that for the noise signal is σ_n^2 . The test statistic D of the energy detection method for M samples is given by

$$D = \frac{1}{M} \sum_{k=1}^M x_k^2 \quad (45.2)$$

D modeled by Gaussian distribution as

$$D \sim \begin{cases} N(M \sigma_n^2, 2M \sigma_n^4), \text{ under } H_0 \\ N(M(\sigma_s^2 + \sigma_n^2), 2M(\sigma_s^2 + \sigma_n^2)^2), \text{ under } H_1 \end{cases} \quad (45.3)$$

By setting decision threshold λ_{ED} , probability of detection P_d^{ED} , and the probability of false alarm P_{fa}^{ED}

$$P_d^{ED} = Q\left(\frac{\lambda_{ED} - M(\sigma_s^2 + \sigma_n^2)}{2M(\sigma_s^2 + \sigma_n^2)^2}\right) \quad (45.4)$$

$$P_{fa}^{ED} = Q\left(\frac{\lambda_{ED} - M \sigma_n^2}{2M \sigma_n^4}\right) \quad (45.5)$$

where Q is the Q -function.

45.3.2 Cyclic Prefix Autocorrelation Detection

The autocorrelation-based detector relies on one-dimensional autocorrelation property that the autocorrelation of the sum of two completely uncorrelated functions (the cross-correlation is zero for all τ) is the sum of the autocorrelations of each function separately. The length of data symbol is T_d . Let H_0 be the null hypothesis, i.e., there is no primary user transmission present, and H_1 be the alternate hypothesis, i.e., primary user is active. The autocorrelation function can be estimated as

$$\hat{R}(\tau) = \frac{1}{M} \sum x(t)x^*(t + \tau) \quad (45.6)$$

where $x(t)$ is the received signal and M is the number of autocorrelation functions. The signal $x(t)$ can be expressed as

$$x(t) = s(t) + n(t) \quad (45.7)$$

where $s(t)$ is the transmitted signal and $n(t)$ is the additive white Gaussian noise (AWGN). It is assumed that $s(t)$ is Gaussian distributed

$$s(t) \sim N_c(0, \sigma_s^2) \quad (45.8)$$

$$n(t) \sim N_c(0, \sigma_n^2) \quad (45.9)$$

and

$$x(t) \sim N_c(0, \sigma_s^2 + \sigma_n^2) \quad (45.10)$$

N_c denotes complex Gaussian distribution. The transmitted data, \mathbf{r} is unknown, and let \mathbf{r} be a random variable. Under H_0 , the mean and variance of \mathbf{r} are given by

$$E[\mathbf{r}] = 0 \quad (45.11)$$

$$\text{Var}(\mathbf{r}) = \frac{(\sigma_n^2)^2}{M} \quad (45.12)$$

Under H_1 , the mean and variance of \mathbf{r} are given by

$$E[\mathbf{r}] = \mu\sigma_s^2 \quad (45.13)$$

$$\text{Var}(\mathbf{r}) = \frac{(\sigma_s^2 + \sigma_n^2)^2 + 2\mu\sigma_s^2}{M} \quad (45.14)$$

Let $\mathbf{r}\mathbf{r}$ be $\text{real}(\mathbf{r})$ used as the test statistic for detection in the low-SNR regime ($\sigma_n^2 \gg \sigma_s^2$).

$$\sigma^2 = \frac{(\sigma_s^2 + \sigma_n^2)^2}{2M} \quad (45.15)$$

Denoting λAC as the statistic threshold, the probability of detection P_d^{AC} and the probability of false alarm P_{fa}^{AC} can be derived

$$P_d^{\text{AC}} = P\left(\mathbf{r}\mathbf{r} > \lambda\text{AC} | H_1 = \frac{1}{2} \text{erfc}\left(\frac{\lambda\text{AC} - \mu\sigma_s^2}{\sqrt{2\sigma}}\right)\right) \quad (45.16)$$

$$P_{fa}^{\text{AC}} = P\left(\mathbf{r}\mathbf{r} > \lambda\text{AC} | H_0 = \frac{1}{2} \text{erfc}\left(\frac{\lambda\text{AC}}{\sqrt{2\sigma}}\right)\right) \quad (45.17)$$

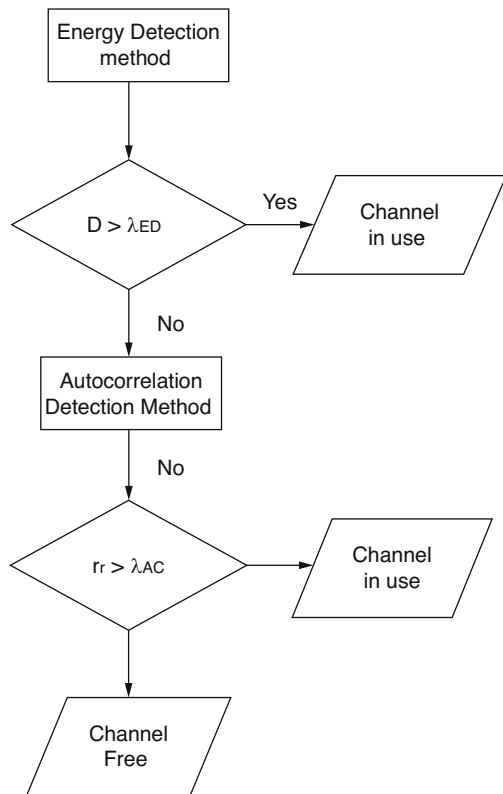
where $\text{erfc}(\cdot)$ is the complementary error function. This method requires prior knowledge of the symbol time. Also, if the prior information of T_c is known, then the performance of autocorrelation detection can be better. Even if symbol time is unknown, we can calculate $\hat{R}(\tau)$ for all values of $T_d = 1, 2, \dots, M$. In [8], it is shown that a priori knowledge of T_d has performance degradation of detection probability.

45.3.3 Fusion Scheme

A fusion of energy detection (ED) and cyclic prefix autocorrelation detection (CPAD) through a cascaded architecture is proposed. As shown in Fig. 45.2, a cascaded architecture has two stages: the energy detector and the autocorrelation-based detector. In the first stage, the channel is sensed using the energy detector, if the result of the energy detection shows that the channel is in use; otherwise the sensing result is decided by the autocorrelation-based detector.

In the first stage, computation is done by the energy detector to calculate decision metric D . If decision metric D is greater than threshold λ_{ED} , the channel is declared to be in use. Otherwise, the received signal is analyzed by the second-stage sensing using autocorrelation-based detection. If the resulting decision metric r_r is greater than the threshold λ_{AC} , the channel is declared to be busy; otherwise, it is declared to be free for transmission. Energy detection is used in the first stage since the energy detector has simpler computation and good performance at high SNRs. At lower SNRs, energy detection becomes unreliable, and we can rely on the autocorrelation-based detector.

Fig. 45.2 Flowchart of the cascaded architecture



In the cascaded architecture, the same number of M samples for both the energy detector and the autocorrelation-based detector to find its detection probability improvement without adding sensing time is used. Another advantage of the cascaded architecture is its energy efficiency.

By substituting P_d^{ED} , P_{fa}^{ED} , P_d^{AC} , P_{fa}^{AC} from Sect. 45.3, we can get the following equation for P_d^{CA} and P_{fa}^{CA} :

$$P_d^{CA} = Q\left(\frac{\lambda ED - M(\sigma_s^2 + \sigma_n^2)}{2M(\sigma_s^2 + \sigma_n^2)^2}\right) + \left[\left(1 - Q\left(\frac{\lambda ED - M(\sigma_s^2 + \sigma_n^2)}{2M(\sigma_s^2 + \sigma_n^2)^2}\right)\right) \times \frac{1}{2} \operatorname{erfc}\left(\frac{\lambda AC - \mu\sigma_s^2}{\sqrt{2}\sigma}\right)\right] \quad (45.18)$$

$$P_{fa}^{CA} = Q\left(\frac{\lambda ED - M\sigma_n^2}{2M\sigma_n^4}\right) + \left[\left(1 - Q\left(\frac{\lambda ED - M\sigma_n^2}{2M\sigma_n^4}\right)\right) \times \frac{1}{2} \operatorname{erfc}\left(\frac{\lambda AC}{\sqrt{2}\sigma}\right)\right] \quad (45.19)$$

According to the cascaded method, the channel is declared as occupied at the time when the energy detection method results in channel in use and also the autocorrelation-based detection results as channel in use. The channel is declared free only when autocorrelation-based detection result is channel idle. Based on this, P_d^{CA} and P_{fa}^{CA} are derived.

45.4 Hardware and Simulation Results

Universal Software Radio Peripheral (USRP) NI 2920 software-programmable radio transceivers are designed for teaching and research in wireless communications.

45.4.1 Energy Detection Technique

The energy detector is implemented. In this detection the QAM recovered clusters are unbundled initially. The unbundle function splits a cluster into each of its individual elements, and the function resizes automatically to display outputs for each element in the cluster. Then the elements are formed into an array and indexed.

Index array function returns the element or sub-array of any dimension array at index. This function resizes automatically to display index inputs for each dimension in the array. The cluster elements are fed to array size function, which returns the number of elements in each dimension of array.

The element returned represents the number of samples N that are used for threshold calculation. Energy detection is implemented by finding the absolute of elements of the cluster and squaring it. Then the squared signals are averaged to determine the energy of the signal. This energy is divided by N , number of samples that gives the total energy of the signal. In order to find this total energy continuously for all samples, these procedures are made to run in a “For loop” where it executes the same N times, number of samples. Threshold is calculated by implementing the formula of λ_{ED} in (10) that requires signal-to-noise ratio and probability of false alarm as input. The total energy value is compared with the threshold value every time. If the total energy value is greater than the threshold, (14) satisfies and represents the presence of a primary signal, whereas if the threshold exceeds the total energy, (15) is satisfied representing the absence of a primary signal. Hence, the energy detection technique is implemented for the detection of primary signals.

45.4.2 Autocorrelation-Based Detection Technique

In the autocorrelation-based detection technique, the QAM-recovered clusters are unbundled initially. The unbundle function splits a cluster into each of its individual elements, and the function resizes automatically to display outputs for each element in the cluster. Then the elements are formed into an array and indexed. The index array function returns the element or sub-array of any dimension array at index. This function resizes automatically to display index inputs for each dimension in the array. The cluster elements are fed to array size function, which returns the number of elements in each dimension of array. The elements returned represents the number of samples N that are used for threshold calculation. The autocorrelation function computes the autocorrelation of the input sequence. Hence, the autocorrelation of the samples $r_{xx}(l)$ is computed as that in (18). Total autocorrelation value is calculated by dividing the out autocorrelation with the number of samples N .

Threshold is calculated by implementing the formula of λ_{AC} that requires signal-to-noise ratio and probability of false alarm as input. The total autocorrelation value is compared with the threshold value every time. Whenever, the total autocorrelation value is beyond the threshold, it represents the presence of a primary signal, whereas if the threshold exceeds the total autocorrelation, it represents the absence of a primary signal. Hence, the autocorrelation-based detection technique is implemented for the detection of primary signals.

45.4.3 Performance of Energy Detection Technique and Autocorrelation Detection Technique

Probabilities of detection by the energy detection technique and autocorrelation-based detection technique for various SNR values are plotted with the help of the simulation tool MATLAB. Probabilities of detection of energy detection and autocorrelation-based detection under various SNRs are shown.

As shown in Table 45.1, at low SNRs, energy detection becomes unreliable. But for higher SNRs, probability starts increasing and reaches 1 as SNR increases to 1 dB. It has simpler computation and good performance at high SNRs. From Table 45.2 the probability of autocorrelation-based detection increases at low SNR and becomes constant as SNR grows but with high computation complexity and requires more number of samples is evident (Fig. 45.3).

Table 45.1 Energy detection technique performance. Detection probability for different SNRs $P_{fa}=0.2$ Number of samples = 1000

SNR (dB)	Detection probability
-20 dB to -10 dB	0
-10 dB to 0 dB	0–0.3
0 dB–5 dB	0.3–0.7
>5 dB	1

Table 45.2 Autocorrelation detection technique performance. Detection probability for different SNRs $P_{fa}=0.2$ Number of samples = 1000

SNR (dB)	Detection probability
-20 dB to -10 dB	0.03–0.07
-10 dB to 0 dB	0.07–0.75
0 dB–5 dB	0.75–0.9
>5 dB	1

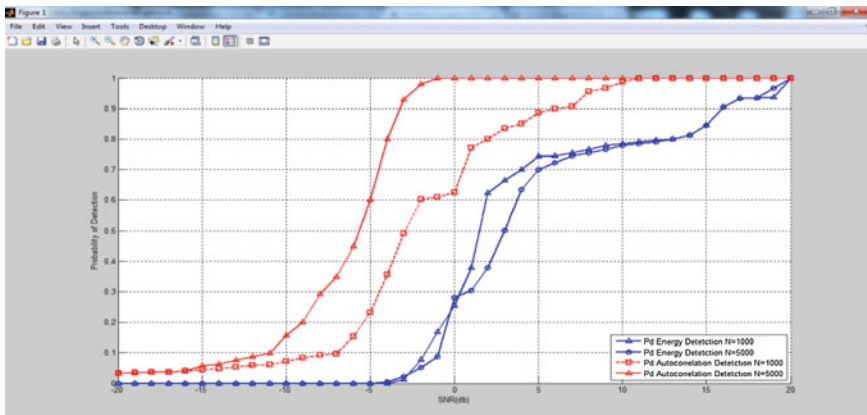


Fig. 45.3 P_d comparison of energy detector and autocorrelation-based detector for $N = 1000, 5000$ samples with varying SNRs under AWGN channel

45.4.4 Cascaded Approach

In the first stage of the block diagram, computation is done by an energy detector using (10) to calculate the decision metric. If the decision metric is greater than the threshold λED , the channel is declared to be busy. Otherwise, the received signal is analyzed by the second-stage sensing using autocorrelation-based detection. In the cascaded block diagram, initially energy detection is done and a comparison is made between energy and threshold. When the decision metric is less than the threshold, then it enters to autocorrelation detection by a case structure.

A case structure contains one or more sub diagrams, or cases, exactly one of which executes when the structure executes or case is selected. The case for selection depends on the decision made by energy detection method. The second stage of cascaded architecture contains autocorrelation-based detection. Case structure is true when the energy detection results in channel free. When the energy detector results as channel busy, then the final result would be busy channel. The probability of detection of the energy detector will be the cascaded output probability of detection. In the channel free case, the autocorrelation block is executed and the test statistic of autocorrelation is compared with that of the threshold. If the decision metric is greater than the threshold, the channel is declared to be free.

The probability of detection after autocorrelation is calculated as the probability of detection of cascaded algorithm. The autocorrelation technique is robust at low SNRs, and they are in the second stage because of computation complexity and requirement of a high number of samples, which will increase sensing time. But in this method, the same number of samples is used without added sensing time. For positive SNRs, a constellation plot is shown. Probability of detection for energy detection increases, and for the cascaded approach, it stables at 1. The performance of cascaded approach is compared with the energy detection and autocorrelation detection using the software simulation tool MATLAB. Probabilities of detection of energy detection, autocorrelation detection, and cascaded approach are plotted, and the results are compared. The results infer that at lower SNRs (below 0 dB), autocorrelation-based detection is better than energy detection, while at higher SNRs (above 0 dB), energy detection is preferred because of its simplicity and good performance (Table 45.3).

Probabilities of detection of energy detection, autocorrelation-based, and cascaded approach for various SNRs are shown in Fig. 45.4. The cascaded approach employs the advantages of energy detection in positive SNR and autocorrelation-based detection in negative SNRs. It is evident from the results that the probability of

Table 45.3 Cascaded approach. Detection Probability for different SNRs
 $P_{fa} = 0.2$ Number of samples = 1000

SNR (dB)	Detection probability
-20 dB to -10 dB	0.0103-0.05
-10 dB to -5 dB	0.0103-0.05
-5 dB to 0 dB	0.28-0.91
0 dB-2 dB	0.91-1
<2 dB	1

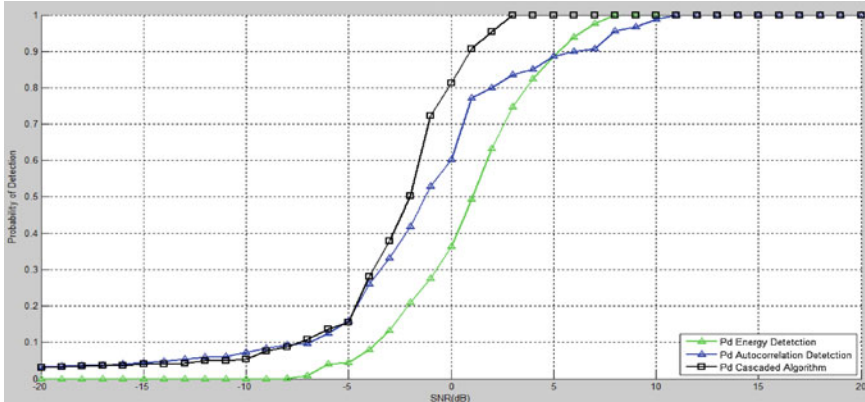


Fig. 45.4 P_d comparison of energy detector and autocorrelation-based detector with the cascaded approach for various SNRs under AWGN channel

detection P_d^{CA} by the cascaded detector is higher than those of the energy detector P_d^{ED} and the autocorrelation-based detector P_d^{AC} .

45.5 Conclusion

In this paper, we proposed a cascaded architecture combining the energy detector and the autocorrelation-based detector. Energy detection involves lower complexity and energy consumption, but unreliable at low SNRs, while the autocorrelation-based detector, which involves more complexity, is more robust at low SNRs. From the analysis, we show that the probability of detection of a cascaded design is always higher than the probability of detection of a single detector. The simulation results confirm that the probability of detection of the cascaded architecture is higher than the probability of detection of the energy detector and the autocorrelation-based detector. The implementation of the cascaded detector allows no additional sensing time. Moreover, the cascaded detector exhibits as an energy-efficient detector by fully utilizing both detectors at low SNRs and only utilizing the energy detector at high SNRs. The proposed architecture for the cascaded algorithm can be verified and improved with more realistic and severe conditions.

References

1. Stevenson CR, Chouinard G, Lei Z, Hu W, Shellhammer SJ, Caldwell W (2009) IEEE 802.22: the first cognitive radio wireless regional area network standard. *IEEE Commun Mag* 47(1):130–138
2. Cabric D, Tkachenko A, Brodersen RW (2006) Spectrum sensing measurements of pilot, energy, and collaborative detection. In: *IEEE military Commun. On proceedings*, IEEE Commun, October, pp 1–7
3. Wild B, Ramchandran K (2006) Detecting primary receivers for cognitive radio applications. In: *1st IEEE international symposium on proceedings. IEEE Commun, New Frontiers Dynamic Spectrum Access Network (DySPAN)*, pp 124–130
4. Dandawate BAV, Giannakis GB (1994) Statistical tests for presence of cyclostationarity. *IEEE Trans Signal Process* 42(9):2355–2369
5. Mitola J III, Maguire GQ Jr (1999) Cognitive radio: making software radios more personal. *Personal Commun IEEE* 6(4):13–18
6. Zeng Y, Liang Y (2009) Spectrum-sensing algorithms for cognitive radio based on statistical co-variances. *IEEE Trans Veh Technol* 58(4):1804–1815
7. Zeng Y, Liang Y (2009) Eigenvalue-based spectrum sensing algorithms for cognitive radio. *IEEE Trans Commun* 57(6):1784–1793
8. Jamali M, Downey J, Wilikins N, Rehm CR, Tipping J (2009) Development of a FPGA-based high speed FFT processor for wideband direction of arrival applications. In: *IEEE Radar conference*, pp 1–2

Chapter 46

Remote Process Monitoring and Control Through IIoT



Swetha R Kumar, Sangavarthini C.S., and L Ashok Kumar

Abstract Internet functionalities and data-based services are introduced in industrial manufacturing processes to innovate the process control based on the Industry 4.0 standards. In this regard, the vision of the IoT extends to closed-loop control. Sensors connect through control algorithms to actuators, with communication over the cloud. Also, the sensor data accumulated in cloud are utilized for process data analytics. These analytics perform predictive control actions by pattern identification, anomaly detection, and outcome prediction from the raw data, as well to trigger control actions through the application of rules. To demonstrate closed-loop control through the IoT, the level of a nonlinear continuous stirred tank reactor plant is continuously monitored and controlled wirelessly using a Wi-Fi-enabled microcontroller CC3200 (Texas Instruments) through an IoT analytics cloud platform. At every time point, the measured values are automatically updated in the cloud. The performance of the IoT-enabled control system is validated by changing various set points. The main implication of this work is to securely and remotely monitor and control various parameters in industry, universally through Internet affiliation.

Keywords Internet of Things · Process data analytics · Nonlinear CSTR

Abbreviations

IIoT Industrial Internet of Things
IoT Internet of Things

S. R. Kumar (✉) · Sangavarthini C. S.
Department of I&CE, PSG College of Technology, Peelamedu/Coimbatore, Tamil Nadu, India
e-mail: srk.ice@psgtech.ac.in

L. Ashok Kumar
Department of Electrical and Electronics Engineering, PSG College of Technology,
Coimbatore, Tamil Nadu, India
e-mail: lak.eee@psgtech.ac.in

CSTR	Continuous stirred tank reactor
LT	Level transmitter
LIC	Level indicator and controller
PV	Process variable
FCE	Final control element
SP	Set point
I/P	Current to pressure
PID	Proportional–integral–derivative
K_P	Proportional gain
K_I	Integral gain
K_D	Derivative gain
MQTT	Message Queuing Telemetry Transport

46.1 Introduction

As technology evolves, industries are looking for automating their manufacturing processes as well as introducing remote monitoring and control systems. Internet functionalities and data-based services can be introduced in these processes to innovate process control based on the Industry 4.0 standards.

The Internet of Things (IoT) is defined as the network of physical objects or “things” embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data. It allows process parameters to be monitored and controlled remotely across existing network infrastructure, allowing more direct integration between the physical world and computer-based systems, and resulting in improved accuracy, efficiency, and cost-effective benefit. Thus, the vision of the IoT extends to closed-loop control. Sensors are connected to actuators through control algorithms, with communication over the cloud [1]. Also, the sensor data accumulated in the cloud shall be utilized for process data analytics. These analytics performs predictive control actions by pattern identification, anomaly detection, and outcome prediction from the raw data, as well triggers control actions through the application of rules [4].

To demonstrate the closed-loop control through the IoT, the level of a nonlinear continuous stirred tank reactor plant is continuously monitored and controlled wirelessly using a Wi-Fi-enabled microcontroller CC3200 (Texas Instruments) through an IoT analytics cloud platform. At every time point, the measured values are automatically updated in the cloud. The performance of the IoT-enabled control system is validated by changing various set points. The main implication of this work is to securely and remotely monitor and control various parameters in industry, universally through Internet affiliation.

46.2 Internet of Things

The Internet of Things (IoT) is the network of physical objects or “things” that can collect and exchange data. The IoT allows variables to be sensed and controlled remotely across existing network infrastructure, giving opportunities for more direct integration between the physical world and computer-based systems. In the current research of the IoT, attention has been focused on wireless sensors, cloud connectivity, Big Data analytics, and mobile apps. From the definitions given above, the vision of the IoT extends to closed-loop control. Sensors connect through control algorithms to final control elements (FCEs), with communication through the Internet [7].

46.2.1 Conventional Closed-Loop Control

A control system is an interconnection of elements forming a system configuration which provides a desired response. For a level control loop, a level transmitter (LT) measures the level in the tank and transmits a signal associated with the level reading to a controller (LIC). The controller then compares the reading to a predetermined value. If the values are equal, the controller then sends a control signal to the device that can bring the tank level back to a lower level through a valve at the bottom of the tank. The valve opens to allow some liquid out of the tank. The controller widely used in industry is a digital PID controller.

In current industrial process control systems, distributed control systems receive data from all remote sensors measuring **process variables** (PVs), compare them with the desired **set points** and derive command functions from control algorithms which are used to control a process through the final control elements.

46.2.2 Closed-Loop Control Through IIoT

When incorporating the IoT in the existing control loop, the conventional controller has to be replaced by a Wi-Fi-enabled processor/controller. This Wi-Fi-enabled device sends sensor data to the cloud and actuates the final control element according to the error. The sensor data accumulated in the cloud shall be utilized for process data analytics [3].

Thus, dedicated control systems can be replaced by the IoT and cloud platforms, which will optimize the performance (such as energy efficiency) for small-scale systems. High-fidelity models could be widely applied for real-time control by cloud-based implementations. Other advantages of cloud-based control are that sophisticated control and optimization algorithms can be implemented, and updated and greater redundancy and fault tolerance could be achieved across critical infrastructures [2].

This scheme has been implemented on a nonlinear continuous stirred tank reactor through a Wi-Fi-enabled microcontroller (CC3200 from Texas Instruments) and ThingSpeak cloud.

46.2.2.1 Process: Nonlinear CSTR

The real-time experimental setup of a CSTR plant is shown in Fig. 46.1. The plant comprises a tank, a water store, a pump, rotameter, a differential pressure transmitter, an electropneumatic converter (I/P converter), a pneumatic control valve with positioner, and a PC. The differential pressure transmitter yield is interfaced with a PC by providing level information to the microcontroller. This microcontroller CC3200 is associated with the USB port of the PC through a USB link.

A pump releases water from the repository, and it moves through the rotameter and electropneumatic positioner. The collection of the fluid in the tank is known as level of the tank. The level of the water in the tank is estimated by methods for the

Fig. 46.1 Experimental setup



differential pressure transmitter and is transmitted as voltage (1–5 V) to the interfacing module and subsequently to the PC. PC is going about as a mistake locator and in addition controller. In the wake of registering the control calculation in the PC, a control signal transmitted to the I/P converter is utilized to activate the electropneumatic positioner to control the valve opening. It controls the stream of the liquid by changing the stem position of the valve opening. For keeping up the level of the procedure tank, the stream is controlled and the rotameter allows the stream of the liquid in the pipeline. For information of water/liquid stream settling rate and yield water/liquid stream settling rate of the CSTR tank, the tank is filled with water until the top or to a liquid capacity of up to 500 mm.

The mathematical model of the CSTR tank system is obtained using the system identification procedure. For fixed input water flow rate and output water flow rate of the CSTR tank, the tank is filled with water up to 500 mm. At each sample time, data from the differential pressure transmitter is collected and fed into the system through an interfacing module.

$$G(S) = \frac{0.0011863S + 0.00184055}{S + 0.00231489} \quad (46.1)$$

The PID controller algorithm is implemented in CC3200 by setting PID gain values as $K_p = 0.6$, $K_i = 0.4$, and $K_d = 0.084$.

46.2.2.2 Controller: CC3200

Created for the Internet of Things (IoT), the SimpleLink Wi-Fi CC3200 device is a wireless MCU that integrates a high-performance ARM Cortex – M4 MCU, allowing end users to develop an entire application with a single IC [9].

The CC3200 Launch Pad is a low-cost evaluation platform for ARM Cortex M4 based microcontrollers. The Launch Pad design highlights the CC3200 Internet-on-a-chip solution and Wi-Fi capabilities.

46.2.2.3 Software Requirements: Energia, ThingSpeak Cloud, and MQTT

In the proposed closed-loop control, Energia is utilized for coding controller algorithms, pushing data on to the cloud, and sending control signals. The ThingSpeak cloud platform is utilized for storing and plotting sensor and controller data. The MQTT protocol is a publish-and-subscribe protocol utilized to change controller parameters' set point remotely through cloud.

46.2.2.4 Implementation of IoT-Based Closed-Loop Control

To implement this control scheme, the followed steps are performed:

- Step 1: Initialize and declare variables like sensor variable, control variable, PID parameter variables, and set point value.
- Step 2: Set the baud rate at 9600 bits per second in Energia.
- Step 3: Receive analog sensor signals to CC3200. CC3200 accepts analog sensor values from 0 to 1.46 V. Thus, sensor voltage (1–5 V) has to be scaled to from 0 to 1.46 V.
- Step 4: Convert voltage to the corresponding level in mm and push the data onto ThingSpeak cloud.
- Step 5: Compute the control signal (1–5 V) using a PID algorithm based on the error value.
- Step 6: The control signal is transmitted to the I/P converter of the CSTR setup for the control valve to take necessary control actions.
- Step 7: The valve position (%) is pushed onto the cloud for monitoring purpose.
- Step 8: The MQTT dash app is utilized to change the controller parameters and set point.
- Step 9: The process parameters values are updated at a constant sampling interval for the control engineer to monitor remotely.

The controlled output and manipulated output is monitored using ThingSpeak cloud and controlled in a graphical field label view by changing the set point and PID parameters using the MQTT dash application in the mobile phone (Fig. 46.2),

The above plot shows the simulation results for level set point = 300 mm and the controlled output from the microcontroller changes with the change in manipulated process variable. It is inferred from the above plot that the controlled output varies with respect to time and error values reduce until the set point is reached (Fig. 46.3).

The above plot shows the valve position output for set point = 300 from the microcontroller changes with the change in manipulated process variable. It is inferred from the above plot that the valve position output varies in percentage and the valve closes on approaching the set point. The valve is completely closed when a set point of 300 mm is attained. These plots are monitored in either a mobile app or a PC as Field 1 Chart and Field 2 Chart, respectively.

Fig. 46.2 Process output for set point = 300 mm

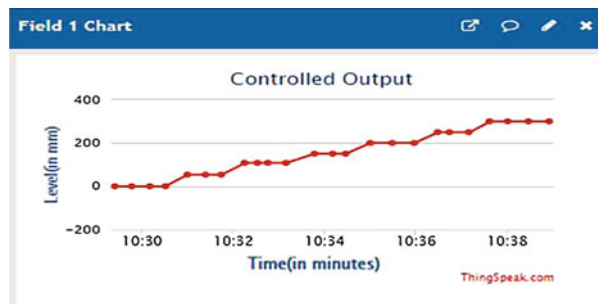


Fig. 46.3 Manipulated variable output for set point = 300 mm

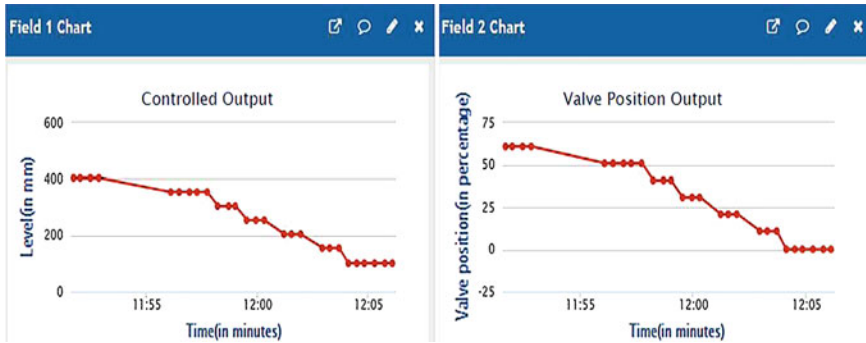
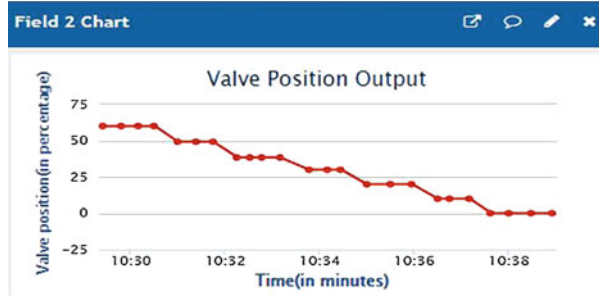


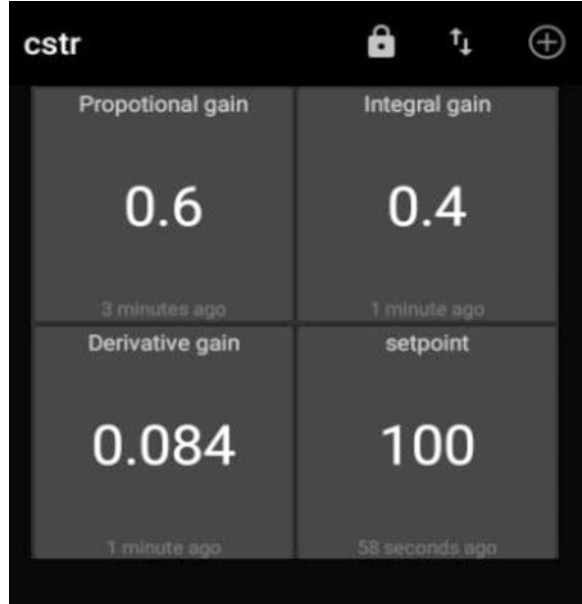
Fig. 46.4 Process output and control input for set point = 100 mm

Similar studies are carried out for a 100 mm set point, and plots are obtained as shown in Fig. 46.4.

To change the controller parameters and desired set point, an MQTT protocol-based app is used. Fig. 46.5 shows the screenshot of the MQTT dash app. The variables initialized in Energia software such as set point, proportional gain, integral gain, and derivative gain are published and subscribed as parameters in the MQTT dash application in the mobile phone. The changes made in the MQTT dash app are immediately reflected in CC3200. According to the changes in the parameters in the MQTT dash app, the controlled output and valve position output are changed and viewed in the ThingSpeak cloud. This MQTT protocol serves as a communication medium by using the topic (subject) of the message to determine which message goes to which client (subscriber).

Thus, a private web server, ThingSpeak, gathers and stores sensor information in the cloud by receiving the analog signals from the microcontroller CC3200. The control signals are generated for the real-time parameters of the continuous stirred tank reactor. The real-time parameters received from CSTR are tuned by the PID controller algorithm implemented in the microcontroller CC3200-equipped Energia software. The graphical trend chart and information logging data about the controlled output and valve position output is achieved by the validating regular change of set points. They are monitored and viewed through either a mobile application or a

Fig. 46.5 Screenshot of the MQTT dash app



personal computer. The web server of ThingSpeak cloud used here is observed to be more secure, and the cyber physical system is globally accepted elsewhere.

46.3 Benefits in IoT-Based Control

The benefits of incorporating the IoT in control loop are as follows [8]:

- (i) Systems that are not physically connected or collocated could be coordinated in real time.
- (ii) Optimized performance (such as energy efficiency) could be achieved for small-scale systems that cannot afford dedicated control systems.
- (iii) High-fidelity models could be widely applied for real-time control via cloud-based implementations.
- (iv) Global networks of sensors and actuators could be implemented and coupled with sophisticated control and optimization algorithms.
- (v) Greater redundancy and fault tolerance could be achieved across critical infrastructures.

46.4 IoT for Process Data Analytics

As many devices are connected to the IoT, the amount of data associated with and produced by IoT devices, including device statuses, metadata, and sensor readings, is growing exponentially. Managing and making meaningful interpretation of these data is essential [5]. Data analytics is implemented on IoT data to create dashboards, statements, visualizations, and alarms, to capture the status and health of connected devices, and to provide visibility for sensor readings [6]. Analytics are used to identify patterns, detect anomalies, and predict outcomes from the data, as well to trigger actions through the application of rules. As a future scope, the obtained level sensor data and control variable data can be utilized for process data analytics.

46.5 Conclusion

In context with Industry 4.0 standards, closed-loop control architecture can be upgraded by utilizing IoT and cloud technologies in their loop. Existing control loop can incorporate the IoT to study the sensor data to perform process data analytics to identify patterns, detect anomalies, and predict outcomes from the data, as well as to trigger actions through the application of rules. Further, the IoT and cloud can perform monitoring and control effectively than the conventional closed-loop scheme as demonstrated in the nonlinear CSTR level loop. This facilitates remote monitoring and control actions.

References

1. Samad T (2016) Control systems and the internet of things [technical activities]. *IEEE Control Syst Mag* 36(1):13–16
2. Ramesh V, Sankaramahalingam M, Bharathy MSD, Aksha R (2017) Remote temperature monitoring and control using IoT, International Conference on Computing Methodologies and Communication (ICCMC), pp 1059–1063
3. Shahzad A, Kim Y, Elgamoudi A (2017) Secure IoT platform for industrial control systems, 2017 international conference on platform technology and service (PlatCon), Busan, pp 1–6
4. Radisavljevic-Gajic V, Park S, Chasaki D (2018) Vulnerabilities of control Systems in Internet of things applications. *IEEE Internet Things J* 5(2):1023–1032
5. <https://www.datascience.com/blog/predictive-analytics-in-industrial-iiot>
6. Ahmed E, Yaqoob I, Hashem I, Khan I, Ahmed AIA, Imran M, Vasilakos A (2017) The role of big data analytics in internet of things. *Comput Netw* 129:459–471
7. <https://www.ibm.com/developerworks/library/iiot-iiot-manage-data/index.html>
8. Khajenasiri I, Estebasari A, Verhelst M, Gielen G (2017) A R view on internet of things solutions for intelligent energy control in buildings for smart city applications. *Energy Proc* 111, Pages 770
9. www.ti.com/lit/ds/symlink/cc3200.pdf

Chapter 47

Case-Based Reasoning (CBR)-Based Smart Indoor Navigation



G. R. Karpagam, K. Eshwar, K. Karthikeyan, and M. Syed Hameed

Abstract In order to improve the efficiency of an urban infrastructure, the concept of smart cities can be used, which is an equal balance between communication technologies and physical devices that connect to them. The recent changes in technology and economy have garnered a significant level of attention in the field of smart cities. The evolution of a city, in terms of both the city infrastructure and the community, can be monitored directly through the help of this technology. One such domain under the smart city infrastructure is the medical domain. On a global scale, hospitals need to take care of factors like pressure with respect to cost and reimbursement, as this can happen when they serve a population with illness. Hence, they need to find out ways to improve their efficiency. By combining the concept of smart city with hospitals, we get smart hospitals. Smart hospitals are being developed with the intention of providing a better value-added service for the common people. It also helps in redesigning and radically thinking about new processes. All of these are enabled by an interconnection of a complete network infrastructure. Further analysis of the smartness in hospitals would lead us to localization. Localization has to heavily depend on sensing devices, such as beacons, RFID tags, etc., despite the use of GPS services. Taking a typical hospital environment into account here, a navigation path should not only provide the shortest path but also keep track of the overall paths the patient has traversed across the hospital. This will help in analyzing the paths of the specific patient and give the respective recommendation measures that need to be taken in order to complete the course of action (in this case, diagnosis and treatment). The motive is a smart indoor navigation system, which learns the user's behavior through previous sensing data.

Keywords Indoor navigation · Case-based reasoning · Ontology

G. R. Karpagam · K. Eshwar (✉) · K. Karthikeyan · M. Syed Hameed
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

Abbreviations

CBR	Case-based reasoning
RFID	Radiofrequency identification
GPS	Global positioning system
IT	Information technology
ICT	Integrated information and communication technologies
IoT	Internet of Things
OWL	Web ontology language
AdOp	Advertisement output
QOp	Query output
ms	Millisecond

47.1 Motivation

The concept of “smartness” in cities can be incorporated into domains that are present, namely, smart buildings, smart water, smart energy and grids, smart public services, smart waste management, and smart mobility. The healthcare sphere’s automation is one of the most tedious tasks to work with. The main aim of the smart hospital system is to create a single IT work environment, which is completely integrated and interconnected.

Medical IT solutions can be broadly classified into telemedicine services, clinical data transmission, patient data’s workflow maintenance and management, and data storage and collection. According to a study, being lost in a hospital environment has been shown to have an increasingly negative impact on the patients and has increased their blood pressure and physical aggression. These solutions, coupled along with the concept of mobility, would bring about the term “navigation” inside a medical system. The main focus of navigating complex healthcare facilities has been on sign and map systems. It has been shown in studies that the state of being lost has several negative impacts on patients and is said to have increased acute stress, blood pressure, physical aggression, and fatigue. Hence, there is an imminent need for the guidance of patients across the indoor environment of a hospital.

47.2 Context of the Work

The context of the work is shown in Appendix Fig. 47.1.

47.3 Related Work

47.3.1 Existing Work in the Area of Smart City

Title – Internet-of-Things-Based Smart Cities: Recent Advances and Challenges

- *Author.* Yasir Mehmood, Farhan Ahmad, Ibrar Yaqoob, Asma Adnane, Muhammad Imran, and Sghaier Guizani.
- *Year.* September 2017.
- *Objective.* A taxonomy to best bring forth a generic overview of the IoT paradigm for smart cities, integrated information and communication technologies (ICT), network types, possible opportunities, and major requirements.
- *Methodology.* The methodologies of this study are as follows: First, a taxonomy of an IoT-based smart city environment is devised. Second, an overview of the major open platforms for smart cities is presented. Third, recent synergies and a number of case studies on various smart city deployments reported by various enterprises are presented. Finally, several IoT-related open research challenges are unearthed.
- *Remarks.* The generic framework of the IoT paradigm combines all the possible communication technologies, provided the fact that we have the required resources to look over the complete framework system, and it has to be self-sustainable. It also should be compatible with the case study we are taking into consideration.

47.3.2 Existing Work in the Area of Smart Indoor Navigation

Title – A State-of-the-Art Survey of Indoor Positioning and Navigation Systems and Technologies

- *Author.* Wilson Sakpere, Michael Adeyeye-Oshin, Nhlanhla B.W. Mlitwa.
- *Year.* December 2017.
- *Objective.* A state-of-the-art survey of indoor positioning and navigation systems and technologies and their use in various scenarios. And analyzing distinct positioning technology metrics such as accuracy, complexity, cost, privacy, scalability, and usability.
- *Methodology.* Using the various positioning algorithms by means of triangulation, trilateration, proximity, and scene analysis and combining the properties of a signal, such as the angle of arrival, the time difference of arrival, and the received signal strength indication, and a general navigation system can be constructed.

- *Remarks.* The paper explains the basic concepts of navigation (outdoor and indoor), and the focus is primarily on the positioning techniques that are involved.

47.3.3 Existing Work in the Area of Case-Based Reasoning

Title – Case-Based Reasoning and User-Generated AI for Real-Time Games [1]

- *Author.* Santiago Ontanon and Ashwin Ram.
- *Year.* 2011.
- *Objective.* This paper focus on a technique that enables end user to create an AI for games without having technical knowledge but by using the case-based reasoning technique.
- *Methodology.* This paper aims to investigate what would make an end user to implement AI for games, which requires a complex strategic reasoning including resource handling and long-time planning under severe real-time constraints and without having complete information.
- *Remarks.* This paper explains the basics of the case-based reasoning and gives a general idea of how to implement that in real-time gaming.

The following studies are also referred for the literature survey: [2–9].

47.4 Lessons Learned

1. There is a definite need for case-based reasoning because it will help the user/patient with the list of suggestions for the next checkup room/treatment room he/she has to visit based on the tests he/she has undergone and the current location of the user/patient.
2. The inherent factors of increase in complexity, due to different technologies that are being introduced for the patient as well as the environment have contributed to the need for indoor navigation (e.g., building automation for energy control and comfort functions for the patients).
3. Any guaranteed solution can be provided only when there is a need for further research to work out the parameters, definitions, and guidelines for the development of the service.
4. The accuracy of obtaining the results depends on the proximity algorithm that is chosen for the system, as well as the signal properties, which has a significant influence on deciding the positioning of the system.

47.5 Objective

The objective of the paper is to design and develop a case-based reasoning (CBR)-based smart indoor navigation system.

47.6 Case-Based Reasoning

Case-based reasoning would refer to using existing experiences of a given task in order to solve a new problem. It does so by being able to remember a previous situation. Using this, it can adapt to the old ones to meet new requests, and we can even critique the solution that we have obtained. The model of CBR is showed in Appendix Fig. 47.2.

47.7 Synergy Between CBR and Smart Navigation System

Let us consider a scenario where A , B , C , D , and E are five different places inside a hospital. Assume that initially in the database we have the following paths: $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$, $D \rightarrow E$. These are the paths that are already in the database. Assume a situation where the user now wants the system to navigate him/her from $A \rightarrow E$. Typically how the system works is that it checks for the direct entry in the table; if that is present, then it just retrieves the data from the database and displays, whereas in this case, the direct entry is not available; this is the place where CBR takes charge; from the previously available data, it tries to form a new path, $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$, $D \rightarrow E$, from which CBR forms a new path $A \rightarrow E$.

And, after forming a new solution that is entered as new data into the table, after insertion, now the table holds the following data, $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow D$, $D \rightarrow E$, $A \rightarrow E$.

For example, if a person met an accident, he will be examined by the general physician in the emergency ward, followed by blood tests and other fluid tests in the pathological lab. Then, he will be taken to the radiology lab where the X-ray and other scans will be taken. He will be taken to the general physician again with the results of the tests and scan reports. Then he will be taken to the operation theatre for surgery. The representation is as follows: *General Physician* \rightarrow *Pathological Lab*, *Pathological Lab* \rightarrow *Radiology Lab*, *Radiology Lab* \rightarrow *General Physician*, *General Physician* \rightarrow *Operation Theatre*.

47.7.1 Representation of Cases in Case-Based Reasoning

While investigating the choice for case representation methods, the following methods were found in the literature. It includes feature vector representation, frame-based representation, object-oriented representation, textual representation, hierarchical representation, and predicate-based representation. Out of the above methods, ontology as case base is chosen since it is based on knowledge base that integrates domain knowledge along with the cases in an ontological structure which enhances semantic reasoning capacity [10].

The matchmaking algorithm is used to find the similarity between the cases. If the new case passed to the system and if the similar case is found in the ontology, that particular case is retrieved and the solution is provided in a shorter span of time. By using ontology, reusability and sharing of cases are enhanced very much. The OWL representation of the cases is shown in Appendix Fig. 47.3. The matchmaking algorithm is based on greedy approach that matches every output concept of the query with one of the concepts of advertisement [11]. It uses the following scheme for degrees of matching. Consider that AdOp is one of the concepts of the outputs of an advertisement and QOp one of the concepts of the outputs of a query.

- *Exact*. If AdOp is an equivalent concept to QOp, then they are labeled as an exact match.
- *Plug in*. If AdOp subsumes QOp, then AdOp can be plugged in place of QOp.
- *Subsumes*. If QOp subsumes AdOp, then service may fulfill the requirements of the request since advertisements provide output in some of the subclasses of the concept defined by QOp.
- *Fail*. If no subsumption relation is found between QOp and AdOp, then it is declared as failure.

47.8 Conceptual Architecture

The Conceptual Architecture of the system is shown in Appendix Fig. 47.4.

- *Registration Agent*. The registration agent would enable the user to register with the hospital servers. This is essential because only an authenticated user can use this indoor navigation facility. The registration agent is shown in Appendix Fig. 47.5.
- *Navigation Agent*. The navigation agent takes care of the entire navigation services inside the hospital. It consists of the mobile navigation module (the front end), the checkpoints that are present within the map, and the overall system. The next steps are typical navigation of the user, where the user is guided

Table 47.1 Observation

Source	Source is known	Source is known	Source is unknown	Source is unknown
Destination	Destination is known	Destination is unknown	Destination is known	Destination is unknown
Case base retrieval time	0.250 ms	0.265 ms	0.259 ms	0.274 ms
Accepted forms	.jpeg, .png formats are supported for the blueprint storage			
Tools	Indoor Atlas		- Open source	
	Android Studio		- Open source	
	Maya		- Open source	
Suggestion based on user experience	Based on the prior inputs that are given by the user, the agent would suggest the next likely destination for the user/patient			

by a path for navigation, and an additional voice command is given to the user, as an assist.

- *Case-Based Retrieval Agent*. The case-based retrieval agent would take care of the new instances of the paths that are provided by the user. This would mean that the agent would learn about potential new paths of the user and help in providing suggestions for the user/patient regarding his/her next visit around the indoors of the hospital. The case-based retrieval agent is shown in Appendix Fig. 47.7.

47.9 Observations

The case ontologies and the matchmaking algorithm are implemented and evaluated using Protégé software. The results of different combinations of observations have been tabulated. The observation is tabulated and shown in Table 47.1.

47.10 Sample Testing

The sample test case is shown in Appendix Fig. 47.6.

47.11 Discussion and Conclusion

The essential summarization of the concept would be as follows: the need for navigation in smart hospitals along with the existential infrastructure of a smart city has led to the proposal of a “Case-Based Indoor Navigation System” which would consist of a case base

along with the navigation module to assist the user in navigating inside the hospital premises. The case-based reasoning method would retrieve, reuse, revise, and retain the cases that are relevant, and this method is specifically used in the context of navigation because of the fact that the result of retrieval of the navigation path is much faster and more efficient when compared to the conventional navigation system. The other advantage of the proposed system is that it is designed to develop incrementally and that the case base can be maintained relatively easier. Hence, the proposed system would be a significant improvement over the conventional navigation system, when it comes to the development of navigation in a smart hospital system, which essentially becomes a part of a smart-city.

Appendix

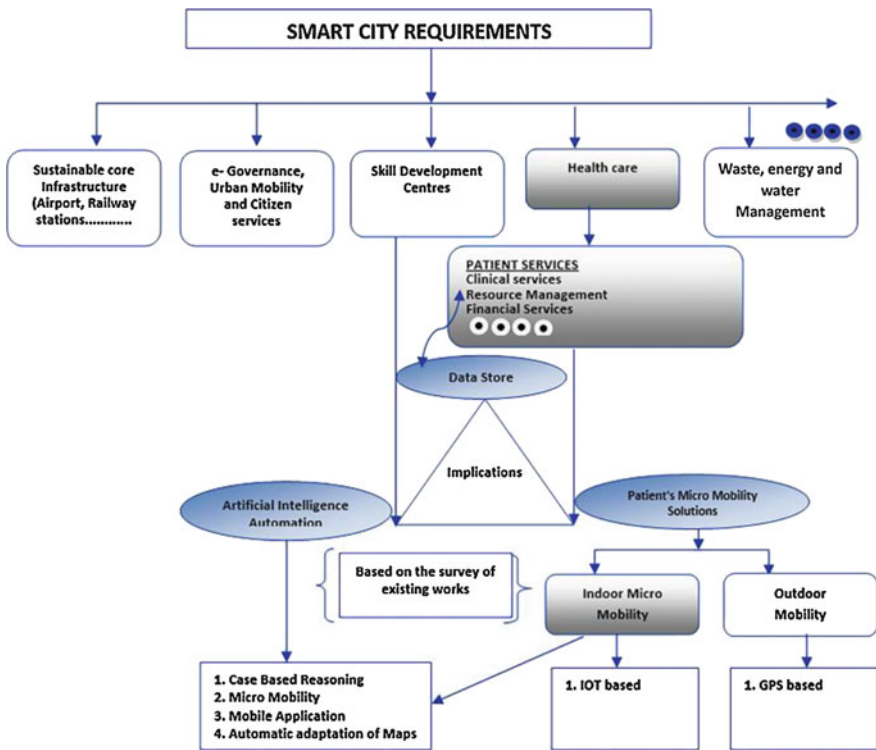


Fig. 47.1 Context of the work

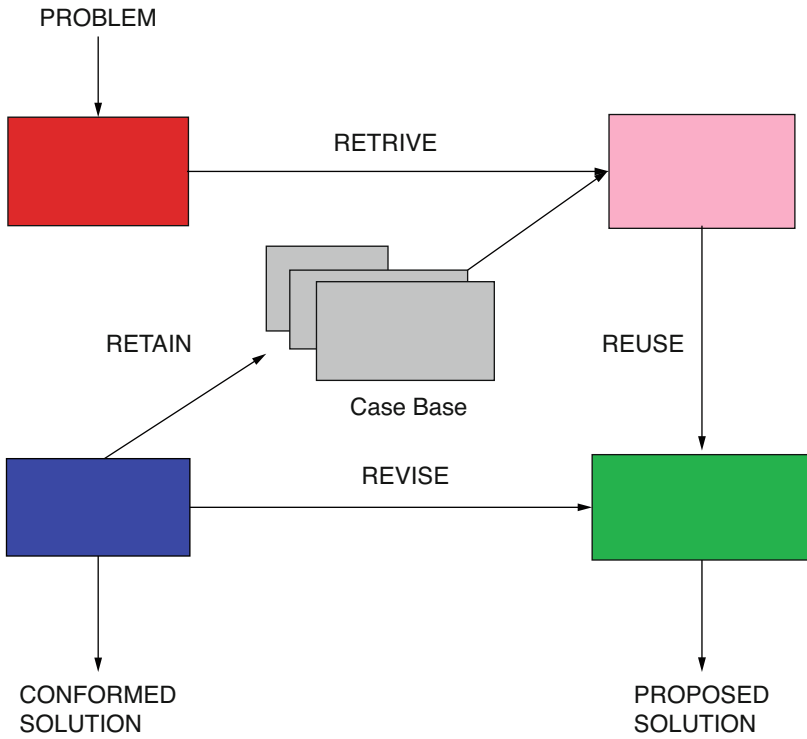


Fig. 47.2 Case-based reasoning

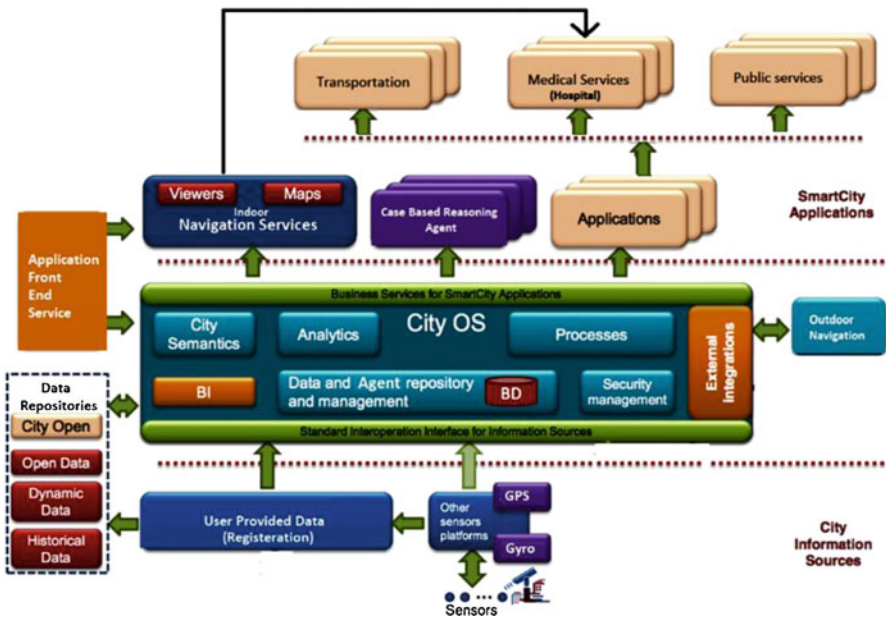


Fig. 47.3 Conceptual architecture

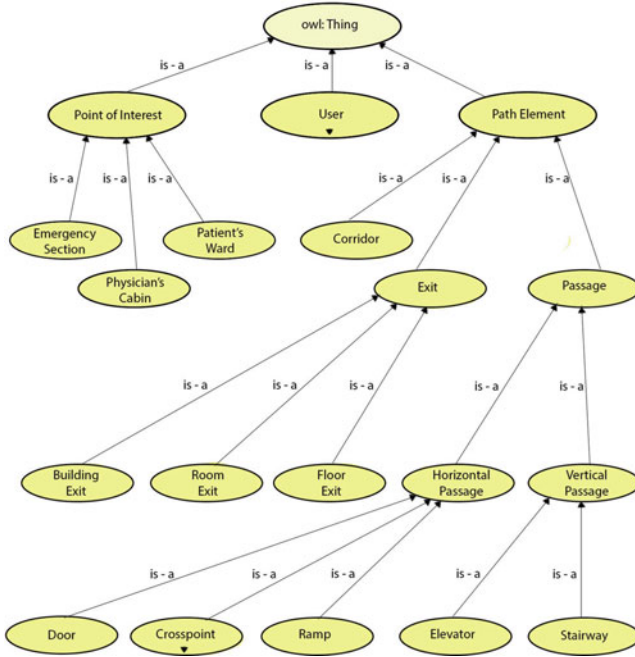


Fig. 47.4 OWL representation of cases

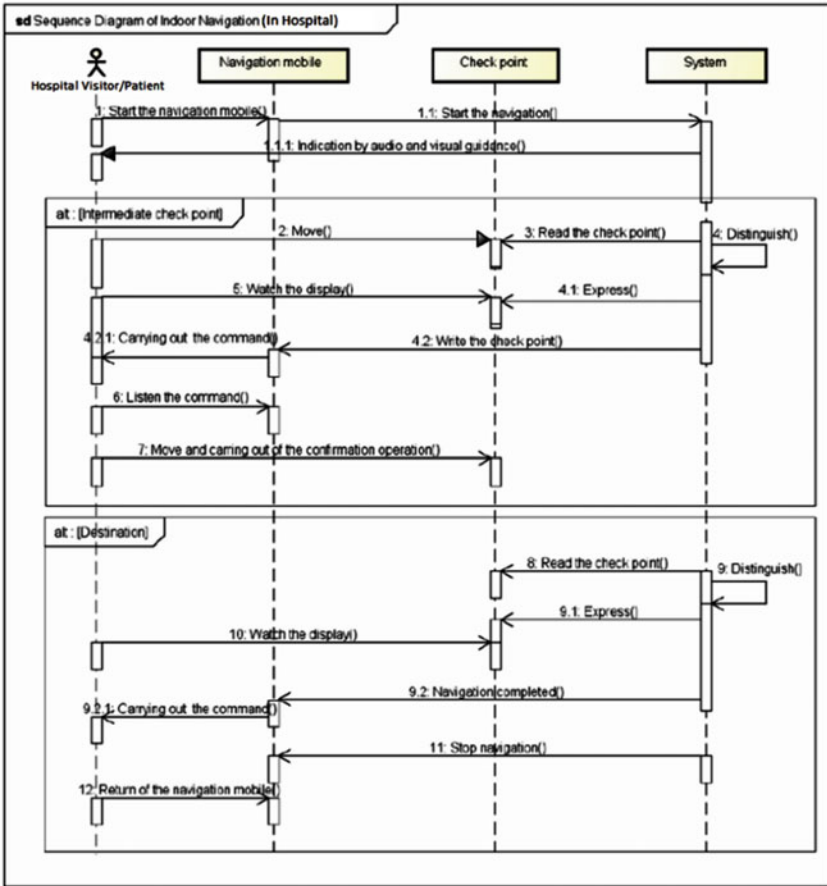


Fig. 47.5 Registration agent

Fig. 47.6 Sample testing

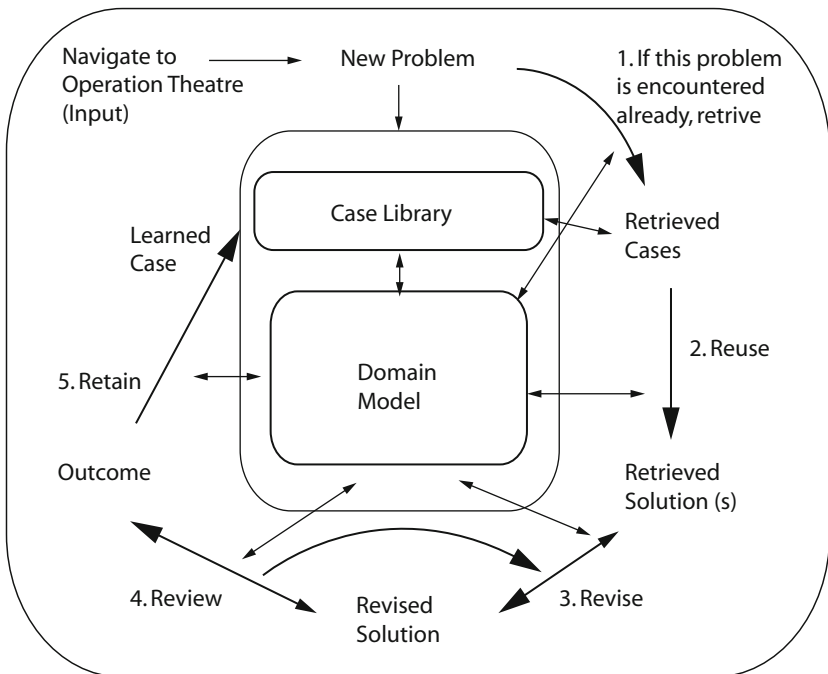
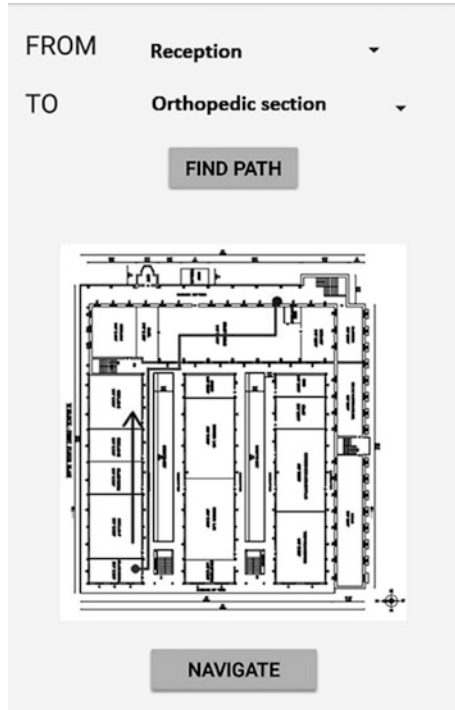


Fig. 47.7 Case-based retrieval agent

References

1. Mehmood Y, Ahmad F, Yaqoob I, Adnane A, Imran M, Guizani S (2017) Internet-of-things based smart cities: recent advances and challenges. *IEEE Comm Mag* 55(9)
2. Trinidad EP, Hinnig MPF, da Costa EM, Marques JS, Bastos RC, Yigitcanla T (2017) Sustainable development of smart cities: a systematic review of the literature. *J Open Innov Technol Market Complex*
3. Komninos N, Mora L (unpublished) Exploring the big picture of smart city research
4. Repko J, DeBroux S (2012) Smart cities literature review and analysis, *IMT 598 Spring 2012: Emerging Trends in Information Technology*
5. Md. Fuzail, Saad J, Khan A (2015) Evaluating the need for smart cities in India, *International Conference on Recent Trends in Engineering Science And Management (ICRTESM)*, At JNU, New Delhi
6. Sakpere W, Adeyeye Oshin M, Mlitwa NBW (2017) A state-of-the-art survey of indoor positioning and navigation systems and technologies. *South African Comp J* 29(3)
7. Van Haute T et al (2016) Performance analysis of multiple indoor positioning systems in a healthcare environment. *Int J Heal Geograp*
8. Gangaputra R (2017) Indoor landmark and indoor wayfinding: the indoor landmark identification issue
9. Ontanon S, Ram A (2011) Case-based reasoning and user-generated artificial intelligence for real-time strategy games. *Artif Intell Comp Games*:103–124
10. El-Sappagh SH, Elmogy M (2015) Case based reasoning: case representation methodologies. *Int J Adv Comp Sci Appl* 6(11)
11. Bellur U, Vadodaria H, Gupta A (2008) Semantic matchmaking algorithms, advances in greedy algorithms, Book edited by Witold Bednorz, Austria

Chapter 48

A Survey on Medical Image Registration Using Deep Learning Techniques



M. C. Shunmuga Priya and L. S. Jayashree

Abstract Image registration is one of the most significant and useful approaches in diagnosing disease by providing complementary information from different medical images. Image registration is a process of overlaying two or more images into a single integrated image. This process is widely used in medical imaging analysis to overlay images obtained from different devices at different time. Traditional methods to geometrically align images are time-consuming, while deep learning techniques are less time-consuming. In recent years, deep learning is a growing technology and has gained many breakthroughs in various image processing problems such as classification, reconstruction, and registration. In particular, convolutional neural networks (CNNs) is one of the most powerful tools in computer vision task. Recently, deep learning techniques are being developed for medical image registration, and image fusion is clearly evidenced from high-quality research. The intention of this survey is to provide perspective about the recent development of registration techniques using machine learning and deep learning techniques.

Keywords Deep learning · Image registration · Reconstruction · Convolutional neural networks

Abbreviations

CT	Computed tomography
MRI	Magnetic resonance imaging
PET	Positron emission tomography
ReLU	Rectified linear unit
LReLU	Leaky ReLu

M. C. S. Priya (✉) · L. S. Jayashree
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

PReLU Parametric ReLu
CNN Convolutional neural network

48.1 Introduction

In medical imaging, the accuracy of diagnosis is highly reliant on image acquisition and image interpretation. Recent image acquisition devices acquire data at high resolution and at a faster rate. In medical diagnosis, physicians require good image interpretation. Image registration helps the physicians to compare pre- and posttreatment images and to analyze disease progression. It is also used in image guided surgery and clinical decision support.

The overall aim of image registration is to find the best optimal alignment between two different images. One of the widely used methods for image registration is the iterative closest point (ICP) algorithm proposed by Besl and Mckay. Because of its simplicity and good performance, it is extensively used. The trimmed ICP algorithm was proposed by Chetverikov and D. Stepanov, which makes use of least trimmed squares in all operations.

The most frequently used medical imaging modalities are computed tomography (CT) scan, magnetic resonance imaging (MRI) scan, ultrasound scan, and positron emission tomography (PET) scans.

48.2 General Workflow of Image Registration (Fig. 48.1)

Feature Detection

Features such as points, lines, contours, and edges from images are detected either manually or automatically. These features are called control points.

Feature Matching

The spatial relationship between features of sensed image and reference image is identified. Various similarity measure techniques such as mutual information and normalized correlation coefficient are used for this purpose.

Transformation Model Estimation

The type and parameters of mapping function are estimated to align sensed image with reference image.

Image Transformation

The image is transformed based on mapping function. The following methods are different affine transformations.

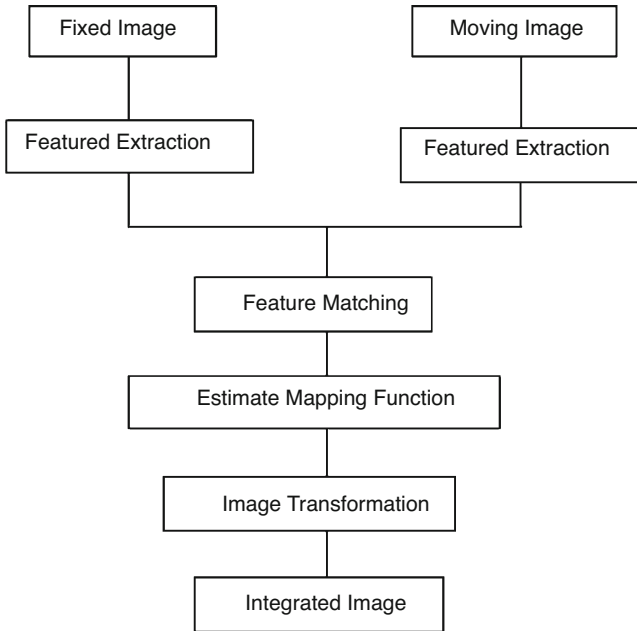


Fig. 48.1 General workflow of image registration

Translation

If X is transformed by t units, then the resulting matrix representation will be as follows:

$$\begin{pmatrix} Y1 \\ Y2 \end{pmatrix} = \begin{pmatrix} x1 \\ x2 \end{pmatrix} + \begin{pmatrix} t1 \\ t2 \end{pmatrix}$$

where x1, x2 is the old point, t1, t2 is the translation value, and y1, y2 is the new point.

Rotation

If a point P (x1, x2) is rotated by an angle θ, then the resultant matrix will be as follows:

$$\begin{pmatrix} Y1 \\ Y2 \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} + \begin{pmatrix} x1 \\ x2 \end{pmatrix}$$

where x1, x2 is the old point, θ is the rotational parameter, and y1, y2 is the new point.

Scaling

Scaling is used to resize an image by either scale up or scale down:

$$\begin{pmatrix} Y1 \\ Y2 \end{pmatrix} = \begin{pmatrix} s1 & 0 \\ 0 & s2 \end{pmatrix} + \begin{pmatrix} x1 \\ x2 \end{pmatrix}$$

where $x1, x2$ is the input point, $s1, s2$ is the scaling factor, and $y1, y2$ is the output point

Shearing

$$\begin{pmatrix} y1 \\ y2 \end{pmatrix} = \begin{pmatrix} a11 & a12 \\ a21 & a22 \end{pmatrix} + \begin{pmatrix} x1 \\ x2 \end{pmatrix} \begin{pmatrix} a13 \\ a23 \end{pmatrix}$$

where $x1, x2$ is the input point, $a11, a12, a13, a21, a22, a23$ is the shearing parameter, and $y1, y2$ is the output point.

48.2.1 Classification of Image Registration (Fig. 48.2)

Image registration is classified on the basis of five factors, namely, dimensionality, modality, subject, transformation model, and interaction. Monomodal registration is used for a single modal such as CT image of a patient between pre- and posttreatment. Multimodal registration is very challenging because it integrates two different imaging modals such as CT and MRI.

Intersubject registration compares images of the same patient taken at different time, while intrasubject registration compares images of different patients. Atlas subject compares an atlas image with the patient's image.

48.3 Overview of Convolutional Neural Networks

After the development of the backpropagation algorithm, the neural network performance improved rapidly. The deeper architecture of CNNs improves the accuracy in medical image analysis.

The input to the networks includes two images. The network consists of a series of convolution layers, followed by pooling layers. These layers process the input images, and the output is a single integrated image.

Training a CNN from scratch with a large amount of data is a great challenge. The accuracy of CNN highly depends on the amount of training data. The higher the amount of training data, the higher the accuracy. But in medical field, the availability of data is insufficient due to privacy policies, so this becomes an issue (Fig. 48.3).

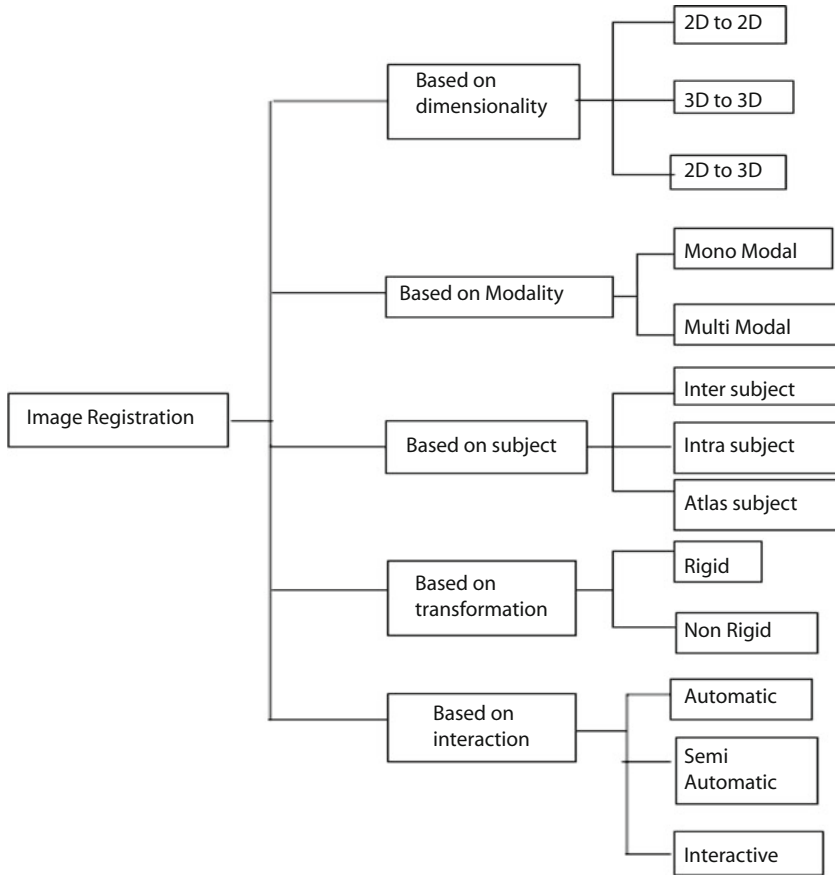


Fig. 48.2 Classification of image registration

48.4 Literature Survey

Miao S et al. [2] proposed a real-time 2D/3D registration that demonstrates how to overcome two major limitations of intensity-based registration: (1) slow computation and (2) slow capture range using CNN regression. It was proved that CNN regression is more accurate and robust than other intensity-based registrations. Bob D de vos et al. [3] described unsupervised deformable image registration with the convolutional neural network. They have proved that DIRNET is as accurate as conventional deformable image registration and improves the computational speed.

R. Liao et al. [1] reviewed the recent advances in registration techniques applied to minimally invasive therapy and methods which reduce operation time and increase accuracy. Guha Balakrishnan et al. [4] proposed an unsupervised learning model for deformable medical image registration which employs deep learning

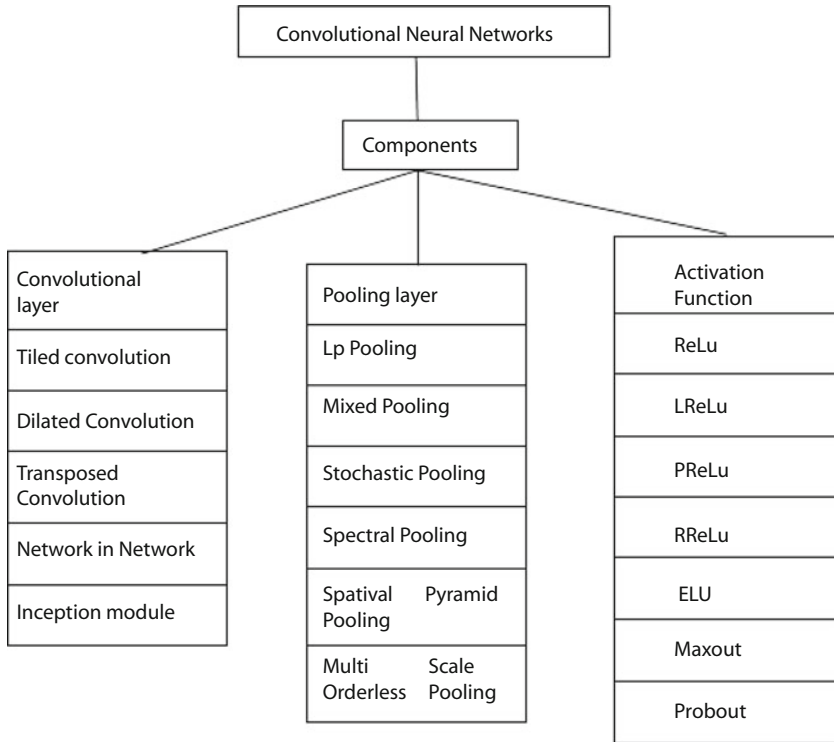


Fig. 48.3 Components of the convolutional neural network

techniques. They have proposed a method called VOXELMORPH and proved that it is faster in analyzing 3D medical images.

B. Rajalingam [5] proposed a novel method of multimodal medical image fusion. They have fused images of different modalities such as CT and MRI using CNN. The main advantage of this method is joint implementing activity level measurement and weight assignment through network training.

Hongming Li and Yong Fan [6] learned spatial transformations between pair of images and maximize image-wise similarity metric. Xiao Yang et al. [7] proposed a deep learning approach to predict 3D deformable registration, and a deep encoder-decoder network is used as the prediction model.

In contrast to other works, H. Sokooti et al. [8] proposed a method for registration called RegNet using CNN, which is tested on 3D Chest CT data. RegNet obtained good results on real-time clinical data. Jingfan et al. [9] proposed a brain image registration method called BIRNet using a fully convolutional neural network with a dual guidance one, that is, coarse guidance using deformation fields and fine guidance using image similarity. BIRNet is a fast, accurate, and easy-to-use method.

48.5 Conclusion

Deep learning Big data has a great impact on the field of medical imaging. This survey presents a deep learning-based image registration algorithm to learn spatial transformation between different images. It covers recent trends with respect to both image registration accuracy and speed. Deep learning methods have great potential in medical image analysis for accurate disease prediction and diagnoses. Various applications such as image-guided surgery can make use of deep learning to facilitate the treatment for patients. Future research in the image registration field can make use of the information from these recent studies to get an overall perspective of the techniques used.

References

1. Liao R, Zhang L, Sun Y, Miao S, Chefd C (2013) A review of recent advances in registration techniques applied to minimally invasive therapy, *Multimedia*. *IEEE Trans Multimedia* 15 (5):983–1000
2. Miao S, Wang J, Liao R (2016) A CNN regression approach for real time 2D/3D registration. *IEEE Trans Med Imag* 35(5). <https://doi.org/10.1109/TMI.2016.2521800>
3. de Vos BD, Berendsen FF, Viergever MA, Staring M, Isgum I (2017) End to end unsupervised deformable image registration with a convolution neural network. Springer, Cham, pp 204–212. <https://doi.org/10.1007/978-3-319-67558-9-24>
4. Balakrishnan G, Zhao A, Sabuncu MR, Gutttag J, Dalca AV, An unsupervised learning model for deformable medical image registration
5. Rajalingam B, Priya R (2018) Multimodal medical image fusion based on deep learning neural network for clinical treatment analysis. *Int J ChemTech Res*
6. Li HM, Fan Y, Non rigid image registration using self supervised fully convolution networks without training data, *IEEE Int Symp Biomed Imag*, 978-1-5386-3636-7
7. Yang X et al (2017) Quicksilver: fast predictive image registration – a deep learning approach. *Neuroimage* 158:378–396
8. Sokooti H, et al. (2017) Non rigid image registration using multi scale 3D convolutional neural networks. In *MICCAI 2017*, Quebec, Canada
9. Jingfan et al. BIRNet: Brain image registration using dual- supervised fully convolutional networks, arXiv:1802.04692. https://doi.org/10.1007/978-3-319-66182-7_27

Chapter 49

Agent-Based Temperature Monitoring System



S. Jaswanth and L. S. Jayashree

Abstract The field of IoT is growing day by day; this growth will lead to wide applications and devices to be interconnected with the help of the Internet. However, there are still many hurdles which have to be rectified before all these will be possible. In this chapter, we have addressed the issue of organising and adapting to the environment where the IoT system is present. To achieve this, we have used the concept of multi-agent systems.

Keywords Internet of Things · Multi-agent System · Organising · Intelligent Environment

Abbreviations

MAS	Multi-agent system
JADE	JAVA agent development framework
GUI	Graphical user interface
ACLMessage	Access control list message
DHT11	Temperature and humidity sensor

49.1 Introduction

In the recent years, a wide variety of applications and devices are being developed which are capable of communicating between themselves, and with the power of machine learning, it is possible for us to use those devices to make decisions [1, 2]. The main advantage of this integration of machine learning and IoT is to

S. Jaswanth (✉) · L. S. Jayashree
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

create an environment capable of adapting to the changing environment without the need for human interference [3].

So our main aim is to develop a system which will monitor the temperature of the environment such as a residential building, and if the temperature increases, that is, a threshold like high temperature, then the cooling system can be turned on and an alert can be sent for the increase in the temperature. All these tasks will be handled by the agent that will control the devices connected to this environment.

49.2 Multi-agent Systems

Multi-agent systems help in creating and handling information that is gathered from sensors placed at various locations. MAS can help in developing systems which will be more robust, help in maintaining the system and be simple [3, 5].

The field of robotics has already contributed a lot to the development of autonomous systems [3]. These methods include the development of robots that learn how to act to the changing environment. All these include various machine learning techniques which will allow the autonomous robots to think on their own [6].

49.3 Characteristics of MAS

49.3.1 Robust

As the multi-agent system is capable of acting to the changing environment on its own, it can handle any errors and environmental changes. This will minimise the need for humans to get involved and correct any errors occurring during the working [3].

49.3.2 Accuracy

An agent-based model should collect all information from the sensory nodes and record it with the time stamp. All these processes are carried out and handled by the autonomous systems, so there will be less errors; the lesser the errors, the more will be the accuracy [1].

49.3.3 Mobility

An MAS can be implemented on any platform and also support cross-platform working which will enhance the feature of mobility. Also, in the future, the same can be mapped to other domains [4].

49.4 Identification of Tool Support

49.4.1 JADE

JADE stands for JAVA Agent Development framework; this framework has been designed to implement the agent-based systems. This framework provides wide features such as message passing, where an ACLMessage can be passed between the agents. This message can be used to make the agents communicate with each other. This feature is the most important feature for creating an agent-based model.

The JADE framework has a GUI interface, so it will be easy for us to control the agents. Also the framework can be installed on a distributed system and can be controlled using a remote GUI (Fig. 49.1).

49.5 Need for an Agent-Based Model

An agent-based model is an autonomous system capable of interacting and performing various actions with the fellow agents present in the environment [3]. All these actions are based on the data which they receive from the sensors. The use of the agent will improve accuracy and reduce the response time of the actuators. In this system, we are building three agents; each agent will perform their own task as specified below:

49.5.1 Monitoring Agent

The main work of this monitoring system is to collect the temperature data from the sensors. Once the temperature exceeds the threshold temperature, which is specified to the monitoring agent, then the monitoring agent sends an ACLMessage to the actuating agent. From this stage, the actuating agent takes control [3].

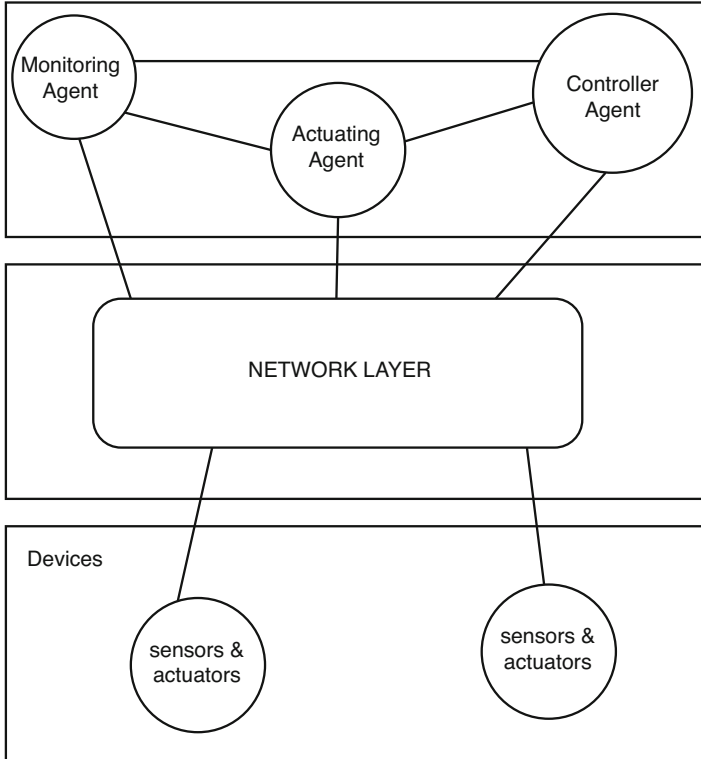


Fig. 49.1 Agent model for monitoring the temperature and performing some actions

49.5.2 Actuating Agent

The actuating agent gets triggered from the message received from the monitoring agent and at times from the controller agent. Once the message is received, the actuating agent triggers the actuators connected to the system [3].

49.5.3 Controller Agent

The controller agent is responsible for making any changes to the system in case of any errors; also, updates can be made to the system, which are needed for the system, by updating the controller agent. The controller agent is responsible for controlling the entire operation of the system [3].

49.5.4 Network Layer

There has to be a communication link for the information and response messages to pass between the actuators and the agents; this network is used for this purpose [4].

49.6 Agent-Based Temperature Monitoring System

As the world is growing fast, there is a need for automation, and the proposed system will help in removing the interference of humans. We have used the JADE framework and have built the Agent Based System, for controlling the sensors and the actuators we have used an Arduino and the sketch was developed in the Arduino IDE.

This board is controlled by the monitoring agent. Temperature samples are received from the sensor at regular intervals. These samples are monitored by monitoring agents. If the temperature exceeds a particular threshold, a warning message is sent along with an ACL message to another agent, which will turn on the cooling fan. Arduino JAVA pack was used to create an interface between the Arduino and JAVA. Once the interface was made, the agent was created.

The system which is created is set up where DHT-11, a temperature and humidity sensor, is used to collect the information about the ambient temperature of the environment. This collected information is sent to the respective agent, and the agent collects this information and checks whether the temperature has crossed the threshold temperature or not. If it crosses, then the monitoring agent sends a message to the controller agent.

The controller agent then triggers the actuating agent, which controls the devices that need to be operated for certain conditions. In this case, we have used a cooling fan setup, which the actuating agent triggers when the temperature reaches the threshold. Also there is a warning for raise in temperature, where another Arduino is connected to a buzzer, and the monitoring agent sends a message to that particular actuating agent, which in turn triggers the buzzer that acts as a warning about the environment.

When a new smart device is added to the environment like a smart kitchen appliance, then an agent can be assigned to that particular appliance and the agent can control it. All these can be done by the agent itself, and this will omit the need for any human intervention. All these automations can help in optimising the power requirement needed for running the devices in home.

49.6.1 Agent Creation Using JADE Framework

Two agents are created, one for taking samples for a regular interval of time and another for issuing a warning. By using the JADE framework, agents are created, and these agents are assigned their specific task to perform. When a new device is being added, then an agent can be assigned, which will communicate with the monitoring agent. The same can be done for multiple devices deployed in the environment. It has to be noted that each agent can communicate between each other only through the asynchronous message (Figs. 49.2, 49.3 and 49.4).

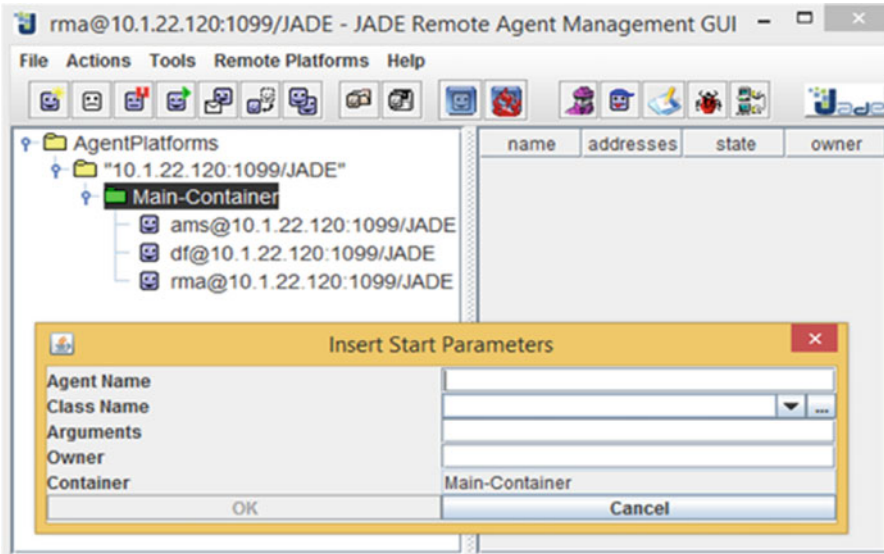


Fig. 49.2 Creation of agents using JADE framework

```
t - jade_proj (run) X
http://workstation2:7778/acc
Oct 15, 2018 9:47:14 AM jade.core.AgentContainerImpl startBootstr
SEVERE: Cannot create agent Agent-1: Class jade_project.Jade_proj
Oct 15, 2018 9:47:14 AM jade.core.AgentContainerImpl joinPlatform
INFO: -----
Agent container Main-Container@10.1.22.120 is ready.
-----
Agent is created
Agent Name Agent1
27.0
27.0
27.0
```

Fig. 49.3 Agent created

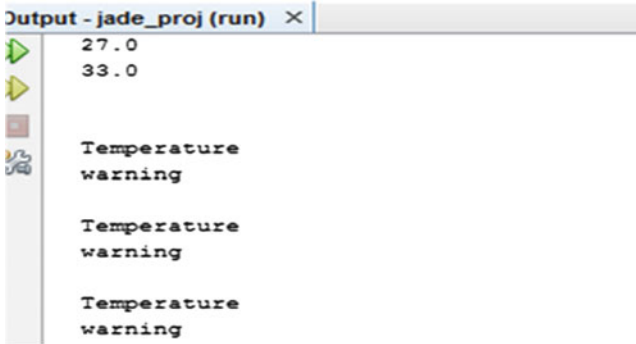


Fig 49.4 Warning generated by an agent

49.7 Conclusion

The growing need for autonomous system has led to the development of many agent-oriented systems. In this chapter, we have addressed a small application that can be developed using an agent-based system and also discussed about the use of the JADE framework, which can be used to develop the same. The reason for us to choose the IoT domain is because the field is rapidly scaling and we need self-to-self-organising agents capable of solving complex problems and always adapting to changing environment.

49.8 Future Work

The entire system was developed based on rule-based approach to estimate the efficiency of the JADE framework. As a future work, we would like to use the machine learning approach where the autonomous system can learn on how to adapt to various changing environments and can give us more effective result. Also, there is only one monitoring system which can monitor only one application, so with the machine learning approach, we can create an agent which can handle multiple applications at the same time.

References

1. Sharma S, Bansal RK (2013) Issues and challenges in wireless sensor networks. In International conference on machine intelligence research and architecture
2. Singh MP, Chopra AK (2011) The internet of things and multiagent systems: decentralized Intelligence in distributed computing, IEEE Conference, pp. 124–138

3. do Nascimento NM, de Lucena CJP (2015) An agent-based framework for self-adaptive and self-organizing internet of things, *Monografias em Ciencia da Computacao*
4. Chen B, Liu W (2010) Mobile agent computing paradigm for building a flexible structural health monitoring sensor network, *Comp Aided Civil Infra Eng*
5. Gubbia J, Buyyab R, Marusic S, Palaniswami M (2013) Internet of things (iot): a vision, architectural elements, and future directions. *Fut Gen Comp Syst* 29:1645–1660
6. Di Marzo G, Karageorgos A, Rana O, Zambonelli F (2004) *Engineering self-organising systems*. Springer, Berlin

Chapter 50

Classification of Phonemes Using EEG



R. Aiswarya Priyanka and G. Sudha Sadasivam

Abstract Artificial speech synthesis can be done using electroencephalography (EEG) and electrocorticography (ECoG) for the brain–computer interface (BCI). This paper focuses on using EEG to classify phonological categories. Although literature is available on the identification and classification of phoneme information in the electroencephalography signals, the classification accuracy of some phonological categories is high, while that of others is too low. Thus, this chapter focuses on identifying the correlation between imagined EEG and audio signals to select the appropriate EEG features. It also identifies the EEG channels that are best suited for imagined speech. Once features are selected, phonemes are classified as vowels and consonants using a support vector machine. Experimental results suggest good accuracy when using 49 features that correlated with audio signals.

Keywords BCI · ECoG · EEG · Phonology · Speech synthesis · Thought to speech conversion · Phoneme extraction · Machine learning · Deep learning

50.1 Introduction

The brain–computer interface (BCI) establishes communication between the brain and an external device. BCI is based on neuroprosthetics. The cortical plasticity of the brain can be used for making the prostheses into a natural sensor or effector channels. For healthy people, the muscles receive commands from the primary motor area through the spinal cord. But this pathway gets interrupted in paralysed people. Thus, motor activity is translated into commands to control the muscles through a computer-based decoder. Neuroprosthetics involves restoration of the impaired function of the nervous system by connecting to an artificial device. The

R. Aiswarya Priyanka · G. S. Sadasivam (✉)
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India
e-mail: gss.cse@psgtech.ac.in

nervous system is connected to the device, and the brain–computer interface is connected to the computer system. Research on using brain signals to control devices or providing imagined speech has gained importance in recent years. The imagined speech is produced by mapping the brain signals to words based on a model to discriminate them. It is of great importance as it can help the disabled people and also patients who are in the locked state.

The brain–computer interface can be invasive, partially invasive or non-invasive. During neurosurgery, invasive brain–computer interfaces are implanted into the brain’s grey matter. Because of this, they provide high resolution, but the scar tissues can build up causing a weak signal. The body can also treat the BCI as a foreign object in the brain and make it non-existent. Partially invasive brain–computer interfaces are also implanted in the same way, but they lie outside the brain’s grey matter. Due to this, they provide better resolution than the previous one. They also possess a low risk of building up of scar tissue. One such example is the electrocorticography. Non-invasive brain–computer interfaces do not require any surgery and can be worn easily. They provide poor resolution because the signals dampen when passing through the skull. They also require prior training to use them for each session. But they can still be used for a wide variety of applications.

The brain’s electrical activity can be monitored electrophysiologically by **electroencephalography**. It is a non-invasive method and uses electrodes to measure the activity. It measures fluctuations in the voltage that is caused by the neurons in the brain due to the ionic current. Due to the introduction of MRI and CT, their use has been decreased for diagnosis purposes. But they are one of the very few techniques that are available mobile. They can be used to diagnose various diseases such as tumours, epilepsy, and stroke. **Electrocorticography** measures the electrical activity of the brain using electrodes from the cerebral cortex. It is a partially invasive method and provides resolution similar to invasive method. **Magnetoencephalography** is used to measure the electrical activity with the use of magnetometers. The brain activity is mapped by recording magnetic fields. It is used to find abnormalities and locations. It provides better spatial resolution than other encephalography methods because of the less distortion. **Functional magnetic resonance imaging** uses blood flow and the activation of the neurons to measure the activity. It does not require surgery and is superior over other methods. It captures changes in blood flow due to the energy used in the brain cells. The electroencephalography and electrocorticography methods are widely used compared to other methods. Although electrocorticography can provide high resolution compared to electroencephalography, it is an invasive method and cannot be used on healthy subjects. Thus, electroencephalography is better when used for research on healthy subjects and do not require surgery.

Section 50.2 provides a short description of the related work on the thought to speech conversion. The Dataset section explains the process of collection of dataset along with its parameters. The Methodology section explains pre-processing, feature selection and classification techniques used. The Experimental Results section highlights results of classification, channel selection and feature selection. Finally, the Conclusion section summarises the information obtained from the results and the future enhancements that can be done.

50.2 Previous Work

Han et al. [5] employed electric and magnetic recordings from seven patients along with participant-specific filters. It was found that recognition varied, but the accuracy of the best two recordings was above 90%. It was concluded that information about the words that are processed is carried over in the brain waves under conditions that were experimented. Independent component analysis was performed to know about the neural processes that occur during the production of speech and vowels using EEG. Neural processes for the production of speech and for mental control were also obtained through the experiment for both auditory response and the speech articulation responses [9]. Neural processes that were task dependent for the production and perception of speech were identified. Ojemann et al. [1] used electrocorticography grids to obtain information from the cortical region. But the grids that were used were miniaturised to obtain more resolution. The patient was made to pronounce audibly each phoneme. Four phonemes were used for the analysis. It showed that different patterns were obtained with respect to the spacing of the electrodes. SVM was used for the classification. This paper was the first to obtain such accuracy for classifying the phonemes. It was also found that the miniature array's sub-regions were distinct for different phonemes. It showed that the processing of the phoneme at the cortex region requires high resolution. Bartels et al. [3] used neurotrophic electrodes for obtaining information from the cortex due to their stability. Localisation of the target in the patients is done by the task of naming using fMRI. The final localisation is done through 3D stereotax during the surgery. Magnetic induction is used to power the electronics that are implanted. Implantation was done on five patients. It was concluded that the signals can be obtained for 4 years from the patients when these electrodes were used. An investigation on whether the information obtained using electroencephalography is enough to determine the thoughts was employed with an experiment after the subjects were made to imagine speaking two syllables [8]. It showed that information was present in the alpha, beta and gamma waves of the EEG signal. For each experimental condition, the filters were computed using the envelopes. The action of the filters on the particular trial's data revealed the condition for such a trial. Schultz et al. [6] used EEG for recognition of unspoken words. The process includes making the subject imagine the words without articulation. The process was done on 21 participants using electroencephalography. For recognition, five words were used, and the Hidden Markov model (HMM)-based classifier model was trained using the words. An accuracy rate of 45.5% was obtained. It was concluded that the temporal information has a high impact on the recognition. Another experiment was done using EEG on three participants [10]. The process used two vowels and no action control state. The nonlinear support vector machine (SVM) was used on the feature vectors obtained. The classification accuracy obtained was between 68% and 70% and proved that vowel imagery can be used in the prostheses controller. Brigham et al. [7] using the electrical waves from the brain explored the electroencephalography's potential for the artificial speech. For the process, seven participants and two syllables were used. The data were pre-processed for noise

removal, and the features were extracted using the autoregressive coefficients. This is done with the help of K-nearest neighbour. It was concluded that the results show that it is feasible to do such classification using electroencephalography. The electrocortical grids are used for obtaining the signals from the face region [2]. The results showed that patterns from the cortical region are associated with the production of phonemes and decoding involves the usage of temporal information. It was stated that a large amount of information can be obtained from the cortex only using ECoG grids and phonemes. Pasley et al. [4] used recordings from the intracranial region. These recordings were from the auditory cortex, which are non-primary. The decoding helped in identifying the words. Although relevant, these studies did not relate the EEG signals with the acoustics or articulation during the speech production. The EEG signals are highly encoded, and this makes it difficult for pre-processing and to select the appropriate features for any task.

50.3 Dataset

The dataset is obtained from 12 participants [11]. The facial and audio information are recorded using Microsoft Kinect. Neuroscan Quik-Cap possesses the 64-channel electrode placement. Then four electrodes are used to control the artefacts from the eye. The SynAmps RT amplifier is used to record the electroencephalography data, and sampling is done at 1 kHz. Seven phonemic prompts are used (/piy/, /iy/, /n/, /tiy/, /uw/, /diy/, /m/) and four words (gnaw, pot, knew, and pat) are considered. Four states are included in each trial: rest state, stimulus state, imagined state, and speaking state. The rest state, which is for 5 s, is used for subjects to relax. This is then followed by the stimulus state. In this state, the prompt on the screen as well as the audio is provided as the stimulus. Then 2 s is needed to move the subjects' articulators for pronunciation. Then it is followed by the imagined state where, without any movement, the subjects are made to imagine speaking the prompts. Then a state of speaking is included where they are made to speak the prompts. The audio and facial features are also recorded during this stage. For each subject, the prompts are provided 12 times each. Thus, each person is provided with trials that are about 132 in total. The subjects were provided rest for every 40 trials. Data of four subjects were removed due to improper recording. For each subject of the 12 subjects, there were 132 trials and for each trial, there were 1197 features. These features include mean, standard deviation, maximum, minimum, median, variance, sum, maximum \pm minimum, spectral entropy, energy, skewness and kurtosis. In addition to these features, the first and second derivatives of these features were also included. Thus, the features obtained in total are about 1197 for each trial, with the 64 channels for each feature.

50.4 Methodology

The overall architecture of the proposed system is shown in Fig. 50.1. The process of phoneme classification involves the following stages.

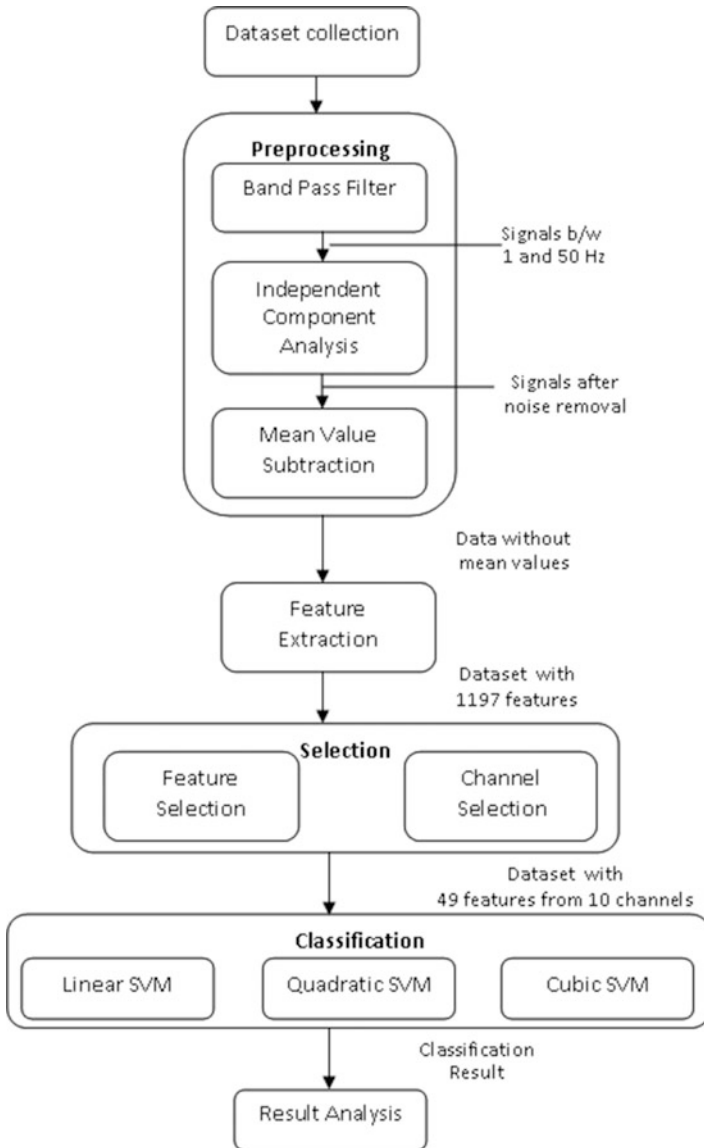


Fig 50.1 Overall system architecture

50.4.1 *Pre-processing*

It includes filtration, blind source separation and mean value subtraction. The band-pass filtration on EEG data is done to extract signals in the frequency of 1–50 Hz [11]. The information in the EEG signal is usually between 1 and 100 Hz. This range comprises the alpha, beta, delta and gamma waves. But they are highly encoded, and extracting information from them is difficult. Thus, to separate the features that are needed, filtration is done.

Blind source separation is a method of unmixing the signals. The original EEG data is highly encoded and is mixed with the noise. Blind source separation is done to unmix those signals and separate the noise (eye movements). Independent component analysis is used for this purpose [11]. In this method, individual components are separated from the multivariate signal. The signal is separated into different components to identify the components that are exposed to high ocular artefacts.

The EEG data between subjects could vary for various reasons. Normalisation mean value subtraction is done to overcome the variations [11].

50.4.2 *Feature Extraction*

Features are extracted from EEG signals. It includes statistical measures such as mean, standard deviation, maximum, minimum, median, variance, sum, maximum \pm minimum, spectral entropy, energy, skewness and kurtosis. In addition to these features, the first and second derivatives of these features were also computed [11]. This gives 1197 features for each channel.

50.4.3 *Feature and Channel Selection*

Dimensionality reduction is done by finding Pearson correlation and selecting the highly correlated features obtained. The correlation coefficients are obtained by finding covariance between the two variables (features or channels), which are then divided by the product of their standard deviations.

- a) **Feature Selection:** The correlation coefficients are found between the EEG features and the audio features. Then the mean of the correlation coefficients of EEG features is found. They are ranked, and the top 49 highly correlated features are obtained. These selected features (mentioned in Sect. 50.5) are then used for classification.

- b) **Channel Selection:** The correlation between the EEG channels and the audio is also found in the same way to obtain the channels (Sect. 50.5) that are highly correlated. After feature selection, the EEG data are passed to the classification phase.

50.4.4 Classification

The EEG signal is classified using support vector machine. The task is to classify the phonemes pair-wise

With feature selection

Without feature selection

With channel selection

Without channel selection

The different classifiers used on the data were

Linear SVM

Quadratic SVM

Cubic SVM

The validation scheme that is used is leave-one-out cross-validation, which removes dependency among the subjects. In this method, the validation is done such that one subject's data is made as test data and remaining as training data for the classification. In the next iteration, test subject's data is added to the training data and some other subject's data is used for testing. This process is repeated until all subjects' data are made as test data and then the result of classification is obtained.

50.5 Experimental Results

EEGLAB in MATLAB is used for EEG signal processing. It includes the pre-processing functions and various other capabilities such as visualisation. EEGLAB supports filtering, independent component analysis, mean value subtraction, removing artefacts, visualising signals, altering the EEG signal, study on EEG data and channel specification.

The 49 features that were selected using the mean of their correlation coefficients include

'Max:W17', 'Max:W4', 'Max:W10', 'Max:W7', 'Max:W16', 'Max:W18', 'Max:W13', 'Max:W3', 'Max:W6', 'Max:W19', 'Max:W8', 'Max:W5', 'Max:W15', 'Max:W11', 'Max:W9', 'Max:W12', 'Max:W14', 'Absmean:W19', 'Max:W2',

Table 50.1 Selected channels

Channel	P3	FC6	CP3	C3	FT8	C4	CP5	C5	CP1	T7
Mean of correlation coefficients	0.382	0.393	0.354	0.391	0.312	0.343	0.374	0.344	0.323	0.311

Table 50.2 Mapping between the features and the most correlated channels

Channel	P3	FC6	CP3	C3	FT8
Features	Absmean:W5	Absmean:W4	Absmean:W17	Max:W19	Max-Min:W9
	Absmean:W3	Max:W13	Max-Min:W17	Max:W11	Max:W14
	Max:W4	Max:W1	Max:W2	Absmax:W3	Absmean:W6
	Max-Min:W18	Absmax:W18	Absmean:W18	Max:W10	Max:W8 Max:W9
	Max:W18	Max-Min:W5	Absmean:W15	Absmean:W13	
Channel	C4	CP5	C5	CP1	T7
Features	Max-Min:W4	Absmean:W8	Max-Min:W19	Max:W7	Absmax:W17
	Max-Min:W6	Max:W15	Max:W5 Max:W16	Absmax:W5	Max:W3
	Max:W12	Absmax:W19	Absmean:W7	Max-Min:W7	Max-Min:W3
	Absmean:W10	Max:W6	Absmax:W4	Absmean:W9	Absmean:W14
	Max:W17	Absmean:W19		Max-Min:W13	Max:W15

‘Absmean:W18’, ‘Max-Min:W19’, ‘Max-Min:W18’, ‘Absmean:W17’, ‘Max-Min:W17’, ‘Absmean:W4’, ‘Max-Min:W3’, ‘Absmax:W18’, ‘Absmean:W5’, ‘Absmax:W3’, ‘Absmean:W3’, ‘Max:W1’, ‘Max-Min:W5’, ‘Absmax:W19’, ‘Absmax:W4’, ‘Max-Min:W4’, ‘Absmean:W14’, ‘Absmean:W9’, ‘Absmean:W10’, ‘Max-Min:W9’, ‘Absmax:W17’, ‘Max-Min:W6’, ‘Max-Min:W7’, ‘Absmean:W15’, ‘Absmean:W6’, ‘Absmean:W7’, ‘Absmean:W13’, ‘Absmax:W5’, ‘Absmean:W8’, ‘Max-Min:W13’.

Table 50.1 presents the channels obtained through correlation and mean of their coefficients. The electrode placement includes the following regions: frontal pole (Fp), central (C), parietal (P), frontal (F), temporal (T) and occipital (O) on the scalp. The letters that code for intermediate electrodes are FC (between F and C), FT (between F and T) and CP (between C and P). The most correlated channels reveal that the temporal regions (T7 and FT8), auditory cortex (CP3 and CP5) and mostly the left part of the brain (C3, CP3, P3, C5, CP1, CP5 and T7) involve in the speech articulation [11] (Table 50.2).

Table 50.3 Vowel versus consonant classification results

Task		Linear SVM	Quadratic SVM	Cubic SVM
With channel selection	With feature selection (24)	53	57	63.8
	With feature selection (49)	54	59	65
	With feature selection (74)	43	49	50.9
	With feature selection (99)	35	38.7	38.9
	Without feature selection	12	12.8	13
Without channel selection	With feature selection (24)	50	57	60
	With feature selection (49)	52	58	60
	With feature selection (74)	47	49	49.9
	With feature selection (99)	32	35	35.1
	Without feature selection	10.7	10.9	11

The classification was done under different conditions as specified before to find the condition that provided high accuracy. Table 50.3 provides the accuracy of different tasks using different classifiers. It can be seen that when 49 features were used the accuracy rate was between 50% and 65% and also when all the 1197 features were used, the accuracy was between 10% and 13%, which is too low.

50.6 Conclusion

Correlated features from EEG data are extracted, and classification between vowels and consonants is done using the support vector machine. Channels correlated with speech signals are extracted. The classification results show that the SVM gives a 60% accuracy rate for the vowel versus consonant tasks with feature selection but gives a below 20% accuracy rate without channel/feature selection. The correlation between the EEG data and the acoustic data helps in identifying the highly correlated channels and features.

References

1. Blakely T, Miller KJ, Rao RPN, Holmes MD, Ojemann JG (2008) Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th annual international conference of the IEEE, pp 4964–4967
2. Kellis S, Miller K, Thomson K, Brown R, House P, Greger B (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neu Eng* 7(5):1–10
3. Bartels J, Andreasen D, Ehirim P, Mao H, Seibert S, Wright EJ, Kennedy P (2008) Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *J Neurosci Method* 174(2):168–176
4. Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex. *PLoS ONE* 10(1):1–13
5. Suppes P, Lu Z-L, Han B (1997) Brain wave recognition of words. *Proc Nat Acad Sci* 94 (26):14965–14969
6. Porbadnigk A, Wester M, Calliess J, Schultz T (2009) EEG-based speech recognition – impact of temporal effects. In: Encarnao P, Veloso A (eds) *Biosignals*. INSTICC Press, Setubal, pp 376–381
7. Brigham K, Kumar BVKV (2010) Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy. In: *Bioinformatics and Biomedical Engineering (iCBBE)*, 2010 4th international conference on, June 2010, pp 1–4
8. D’Zmura M, Deng S, Lappas T, Thorpe S, Srinivasan R (2009) Toward EEG sensing of imagined speech. In: Jacko JA (ed) *Human-computer interaction*. New trends, Lecture notes in computer science, vol 5610. Springer, Berlin/Heidelberg, pp 40–48
9. Callan DE, Callan AM, Honda K, Masaki S (2000) Single-sweep EEG analysis of neural processes underlying perception and production of vowels. *Cogn Brain Res* 10(1-2):173–176
10. DaSalla CS, Kambara H, Sato M, Koike Y (2009) Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Net* 22(9):1334–1339. *Brain-Machine Interface*
11. Zhao S, Rudzicz F (2015) Classifying phonological categories in imagined and articulated speech. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp 992–996

Chapter 51

Attribute Table-Based Multipath Routing Protocol to Improve Network Lifetime in Multi-hop WSN



B. Sherin, M. Senthil Vadivu, and A. Ayub Khan

Abstract Unbalanced energy consumption is an inherent problem in WSNs characterized by multi-hop routing and many-to-one traffic pattern. This uneven energy dissipation can significantly reduce network lifetime. In this paper, we proposed an attribute table-based energy-efficient and QoS multipath routing protocol to improve network performance in the multi-hop network. The attribute table is prepared for each transaction, and the energy level of the node is monitored for every transmission. The optimized pair shortest path algorithm is used to find the shortest path between all pairs of vertexes and reduce the duration of data transmission. After every transmission node, the energy level is noticed by the attribute table for the next process. This efficiently transmits the data with security and minimizes the packet losses.

Keywords Multi-hop routing · Attribute table · QoS · Dynamic route switching

Abbreviations

ATEQMR	Attribute table-based energy-efficient QoS multipath routing
AODV	Ad hoc on-demand distance vector
DSR	Dynamic source routing
WSN	Wireless sensor network
QoS	Quality of service
SHR	Self-healing ring
EQMH	Energy-efficient and QoS-based Multi-path hierarchical routing protocol
MH-RF	Multi-hop radio frequency

B. Sherin (✉) · M. Senthil Vadivu · A. Ayub Khan
Sona College of Technology, Salem, Tamil Nadu, India

51.1 Introduction

Remote systems are opposed to convey using a unified framework; each host goes about as a switch to transmit packets to different nodes. At the point when a source node sends information packets to a goal hub that is not inside the source's transmission run, the containers must be submitted by different nodes. There are regularly distributed networks that generally allow transmission only to a node's nearest neighbors. The nodes in these networks are generally identical, so that mesh nets are also referred to as peer-to-peer nets. Mesh nets can be good models for large-scale networks of wireless sensors that are distributed over a geographic region, for example, personal or vehicle security surveillance systems. An advantage of mesh nets is that, although all nodes may be identical and have the same computing and transmission capabilities, certain nodes can be designated as "group leaders" that take on additional functions. If a group leader is disabled, another node can then take over these duties.

All nodes of the star topology are connected to a single hub node. The hub requires greater message handling, routing, and decision-making capabilities than the other nodes. If a communication link is cut, it affects only one node. In the ring topology, all nodes perform the same function, and there is no leader node. Messages generally travel around the ring in a single direction. The self-healing ring network (SHR) shown has two rings and is more fault tolerant. In the bus topology, messages are broadcast on the bus to all nodes. Each node checks the destination address in the message header and processes the messages addressed to it.

The term "energy efficiency" as used here includes these economic and environmental savings entusing less energy at any time including at times of mates are extrapolations of the results from region peak demand through demand response and peak at program to a national scope. In the routing protocol design of mobile nodes, many issues need to be considered in order to offer many important properties such as scalability, QoS support, security, low power consumption, and so on. So, power consumption is independent of the transmission distance between adjacent nodes. Multipath directing utilizations have numerous ways to transmit information, which can accomplish both loads adjusting and adaptation to internal failure. There are two distinctive multipath routings between the source hub and the sink hub. Exhibited disjoint multipath was addressing in light of neighborhood data, which is a circulated calculation by which stack adjusting can be accomplished. This calculation utilizes an essential course to transmit information.

51.2 Related Works

The wireless sensors in the environment are used to sense, gather, and communicate the information with other sensors; information like pressure, temperature, humidity, and intrusion is received by the sink, which is the basic work done in a wireless

sensor network [1]. This problem is modeled as finding the maximal number of covers, each of which can cover all targets to be monitored [2].

When nodes move slowly or fast within a small network area, AODV and DSR both show good performance of multi-hop data transmission in terms of end-to-end delay, packet delivery rate, and energy consumption [3]. Pairwise synchronization is a common technique for two nodes to synchronize their clocks; a leading protocol that extends it across a larger network is AMTS [4]. An energy-efficient and QoS-based multi-path hierarchical routing protocol (EQMH) in wireless sensor networks provides service differentiation by giving real-time traffic absolute preferential treatment over the non-real-time traffic [5]. It determines the optimal routing path from the source to the destination by favoring the highest remaining battery power, a minimum number of hops, and minimum traffic loads [6]. For a large WSN, network lifetime is inversely proportional to the cluster diameter in terms of a number of hops and the cost is directly proportional to the number of clusters, that is, the number of sinks [7]. The novel fuzzy nondominance shortest path technique considers that network link weight is naturally uncertain due to inaccurate information. It supports a better network QoS with less computational and management cost [8].

51.3 Attribute Table-Based Energy-Efficient QoS Multipath Routing

In this proposed system, we use attribute table-based multipath routing protocol to improve network performance through efficient utilization of the available network resources, If any of the network path is disconnected, the attribute table tracks the next optimized path without any data loss. The attribute route table stores the energy of each node, source to destination possible number of maximum route path, and source to destination duration of communication time for each transaction, in a predefined manner.

51.3.1 Block Diagram

Above the Fig. 51.1 packet which needs to send is collected and sent to Attribute table, in this attribute table the large size packets are divided into multiple packets to reduce the attacks and losses. Then the shortest path algorithm is used to find the shortest way with the highest energy level to transfer the packet. The packet transmission is performed at high speed based on the energy level of the path. On the receiver side, the packet collector is used to collect multiple packets and arrange based on the order. Arranged packets are sent to a receiver node. After completion of transmission, the path energy level is calculated for the next transmission; if the energy level is low, then the path energy level is boosted. The energy level is noticed during the time before and after the data transmission.

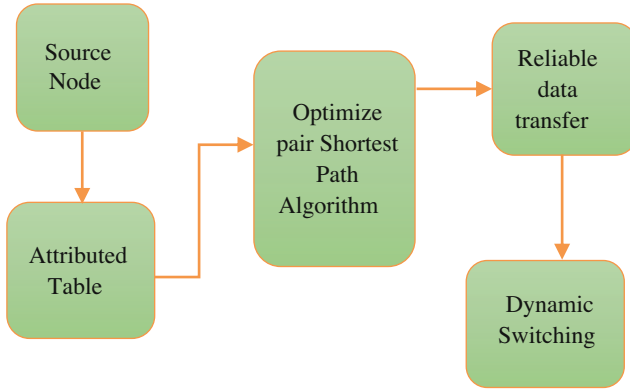
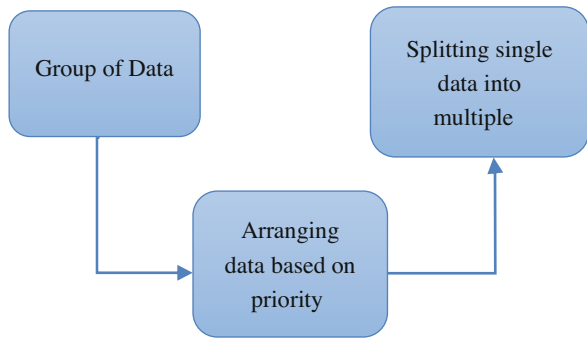


Fig. 51.1 Attribute table-based energy-efficient QoS multipath routing

Fig. 51.2 Block diagram for attribute table process

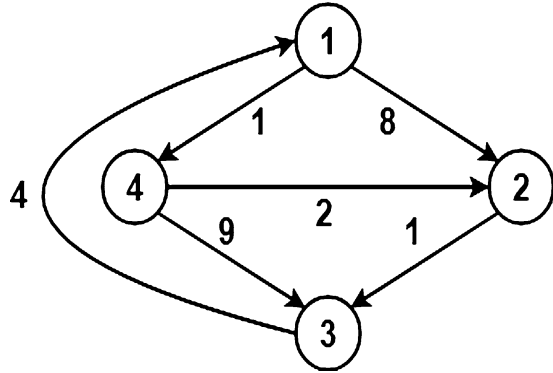


51.3.2 Attribute Table

In this module, an attribute table-based multipath routing protocol is used to transmit the data energy efficiently and reduce loss of data during transmission. In this protocol, the data are arranged and split into multiple data for efficient transmission. When the routing path energy is minimized or any other problem to connection braking that time dynamic route switching protocol is discovered next optimized route path is switch over to the communication. Due to multipath routing, increment in the number of conceivable routes expands the heartiness and throughput of the transmissions. Multipath routing is utilized either for stack adjusting or for consistent quality.

Figure 51.2 shows the attribute table process. In this, “Group of data” is collected and arranged based on the priority, where priority is assigned on the basis of the size of the packet. A large packet is sent first. Another process is splitting large-sized data into multiple small-sized packets and sent through a different path.

Fig. 51.3 Example of shortest path algorithm



51.3.3 Optimized Pair Shortest Path Algorithm

The optimized pair shortest path algorithm is used for solving all pairs shortest path problem, which gives the shortest path between every pair of vertices of the given nodes. The optimized pair shortest path algorithm is an example of dynamic programming. The main advantage of the optimized pair shortest path algorithm is that it is extremely simple and easy to implement. Also it consists of three loops over all nodes. The innermost loop consists of only operations of a constant complexity. Hence, the asymptotic complexity of the optimized pair shortest path algorithm is $O(n^3)$, where n is the number of nodes in the given graph.

As in Fig. 51.3, we can find the shortest path by using the optimized pair shortest path algorithm. The shortest path for node 1 \rightarrow 3 is

$$(1, 3) \Rightarrow (1, 4, 3) = 10(1+9) \text{ (length of node)}$$

$$(1, 3) \Rightarrow (1, 2, 3) = 9(8+1) \text{ (length of node)}$$

$$(1, 3) \Rightarrow (1, 4, 2, 3) = 4(1+2+1) \text{ (length of node)}$$

The values shown above are the length of the node. From this, we can calculate the shortest length to transmit the data from source to destination. The shortest distance for (1, 3) is (1, 4, 2, 3), and the length is 4; based on this, the optimized pair shortest path algorithm is performed for all pairs.

51.3.4 Network Design and Synchronization of Multiple Nodes

This module is developed to create more than 10 nodes placed at a particular distance. A wireless node is placed in the intermediate area. Each node knows its location relative to the sink. The access point has to receive, transmit packets, and then send an acknowledgment to the transmitter. Sensor networks most often have a much more complicated topology than the simple examples, and not all sensor nodes can communicate with each other directly. Thus, multi-hop synchronization is

required, which adds an additional layer of complexity. Clearly, this could be avoided by using an overlay network, which provides virtual, single-hop communication from every sensor node to a single master node.

51.3.5 Energy Consumption

Each sensor node transmits and receives data with fixed transmission and reception power, respectively. So, power consumption is independent of the transmission distance between adjacent nodes. Accordingly, we adopt the following energy model due to calculating the power consumption. A multi-hop network is a multipath network used to transmit data from source to destination through several nodes. Wireless ad hoc networks involve self-directed nodes that cooperate to transport data. Generally, these nodes perform as end systems and routers at the same time. Because of the mobility and scalability of most mobile ad hoc networks, they often exhibit multiple hops for obstacle intervention, range reuse, and energy preservation.

51.3.6 Data Collection Process

In this module, data are collected from different nodes, arranged in order, and sent to the destination. It requires cooperation among a large number of nodes. One example is to continuously monitor an area and report events. Another example is a node that sends out a query about interested data to a number of nodes. Thus, much individual data needs to be collected and extracted to form some higher level information. The procedure of data collection and extraction is called data collection.

51.4 Result and Discussion

The extent of throughput, transport, put-off execution general framework appearance indicate change in orchestration of ordinary and little package release extent and cut packet delay. End delay is figured to keep up a critical distance from the development pantomime structure. Here we have strategies for a typical support that show for decreasing the frame delay and stay away from the action on masterminding, so we have a prevalent result differentiation and conceivable methodology (Figs. 51.4 and 51.5) (Table 51.1).

$$D = (Tr - Ts)$$

Tr – Receive time

Ts – Sent time

Fig. 51.4 Comparison of throughput

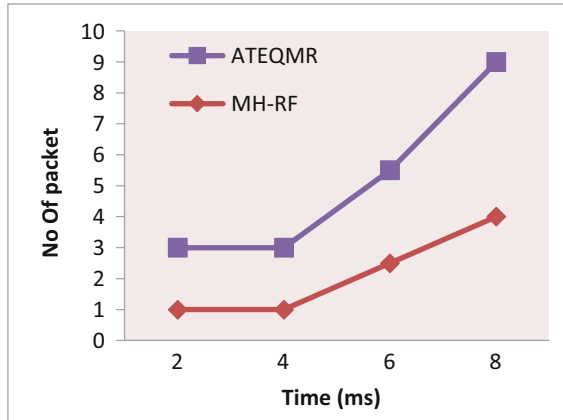


Fig. 51.5 Comparison of the delivery ratio

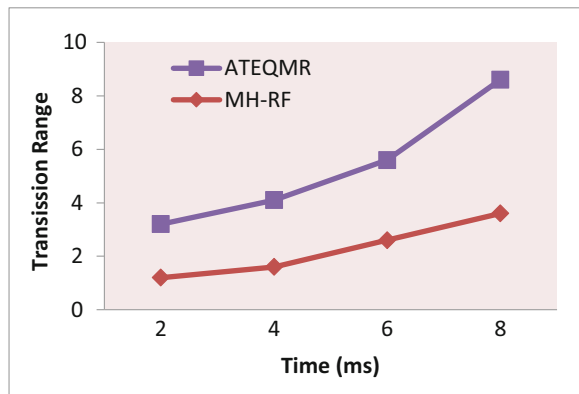


Table 51.1 Comparison of throughput

Time (ms)	No. of packets delivered in MH-RF	No. of packets delivered in ATEQMR
2	1	2
4	1	2
6	2.5	3
8	4	5

51.4.1 Data Delivery Fraction

The packet passed on or after starter place to reason on their framework. The message essentialness is required to transmit or tolerate packages from side to side transmission control or load task and besides the imperativeness, utilization can be constrained in the framework (Table 51.2).

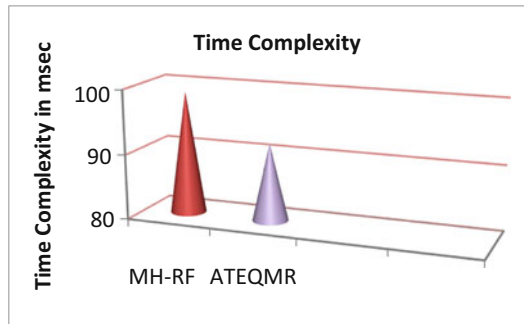
Table 51.2 Comparison of delivery ratio

Time (ms)	Transmission range of MH-RF	Transmission range of ATEQMR
2	1.2	2
4	1.6	2.5
6	2.6	3
8	3.6	5

Table 51.3 Comparison of time complexity

Algorithm	Time complexity (ms)
MH-RF	99
ATEQMR	92

Fig. 51.6 Comparison of time complexity in message authentication process



It is intended that in-between the amount of data documented by the termination state from side to side, the calculation package originates from the starting position on the set of relations.

$$PDF = (Pr/Ps) * 100 \tag{51.1}$$

where Pr is the total data received and Ps is the total data sent on their network (Table 51.3).

Figure 51.6 shows the comparison of time complexity introduced by various methods in location verification, and it also shows clearly that the proposed method has produced less time complexity than others.

51.5 Conclusion

This wireless network uses the attribute table-based energy-efficient and QoS multipath routing protocol to improve the communication data transmission, reduce data loss, and growth for throughput and for efficient data delivery without data loss

using multipath routing protocols. The optimized pair shortest path algorithm is used to send data efficiently for a short distance. During the transmission, if any packet transmission has failed and if the path is attacked, then it will be sent by another path. After every transmission, the node energy level is noticed by the attribute table for the next transmission. If the energy level is reduced, then it will boost for the next transmission; this will reduce path loss during transmission. The result shows that 90% of energy is efficient to transmit data with security and reduce packet losses. In future works, module ideas are developed and implemented, and simulated results will be given based on the required throughput.

References

1. Roshan K, Sharma KR (2018) Efficient routing mechanisms used for network lifetime enhancement in wireless sensor network, *IEEE*:256-260; Lin C-C, Deng D-J, Shu L, Wang K, Wang S-B, Tsai I-H (2016) On lifetime enhancement of dynamic wireless sensor networks with energy-harvesting sensors, *IEEE*:1-3
2. Zhang J, Sun Z (2016) Assessing multi-hop performance of reactive routing protocols in wireless sensor networks. *IEEE*:444-449
3. Gordon S, Kaemarungsi K (2016) Analysis of adaptive multi-hop time synchronization in large wireless sensor networks, Nattaphol Sangjumba. *IEEE*:79-84
4. Dehnavi M, Mazaheri MR, Homayounfar B, Mazinani SM (2012) Energy efficient and QoS based multi-path hierarchical routing protocol in WSNs. *IEEE*:414-418
5. AlShawi IS, Yan L, Pan W, Luo B (2012) Lifetime enhancement in wireless sensor networks using fuzzy approach and a-star algorithm. *IEEE*:1-6
6. Chatterjee P, Das N (2015) Multiple sink deployment in multi-hop wireless sensor networks to enhance lifetime. *IEEE*:48-54
7. An J, Pangalos P, Aghvami AH (2013) Novel fuzzy non-dominance shortest path routing and path ordering for QoS aware routing. *IEEE*:3381-3385
8. Lo C-C, Hu Y-C (2012) Edmas – a latency minimization protocol with low duty-cycle for wireless sensor networks

Chapter 52

Application of Subjective and Objective Integrated Weightage (SOIW) Method for Decision-Making (MADM) in Distribution System



Sachin Gorakh Kamble, Kinhal Vadirajacharya,
and Udaykumar Vasudeo Patil

Abstract The term smart grid (SG) has been used by many government bodies and researchers, which refers to the new trend in the power industry to modernize and automate the existing power system. SG must utilize the assets optimally by making use of information, like equipment capacity; voltage drop; radial network structure; minimizing investment, operating costs, and energy loss; and reliability indices. One way to achieve this is to reroute or reconfigure the distribution system. The distribution system is reconfigured to choose a switching combination of branches of the system that optimize certain performance parameters of power supply while satisfying some specified constraints. In this chapter, the subjective and objective integrated weightage (SOIW) multiple attribute decision-making (MADM) method is proposed for finding the compromised best configuration and comparing it with other methods such as WSM, WPM, and TOPSIS. An example of the distribution system is presented in this chapter to demonstrate the validity and effectiveness of the method.

Keywords Distribution system Reconfiguration · Multiple attribute decision-making · SOIW method, TOPSIS

S. G. Kamble (✉) · K. Vadirajacharya
Dr. Babasaheb Ambedkar Technological University, Lonere, Maharashtra, India
e-mail: sachinkamble80@rediffmail.com; kvadirajacharya@dbatu.ac.in

U. V. Patil
Government College of Engineering, Karad, Maharashtra, India

Abbreviations

AENS	Average energy not supplied
AHP	Analytic hierarchy process
CAIDI	Customer average interruption duration index
CAIFI	Customer average interruption frequency index
DM	Decision maker
DS	Distribution system
ELECTRE	Elimination and choice translating reality
MADM	Multiple attribute decision-making
Pi	Performance Index
SAIDI	System Average Interruption Duration Index
SAIFI	System Average Interruption Frequency Index
SG	Smart grid
SOIW	Subjective and objective integrated Weightage
TOPSIS	Technique for order preference by similarity to ideal solution
WPM	Weighted product method
WSM	Weighted sum method

52.1 Introduction

The electric power system is a vital part of modern developed societies. The electric power system is mainly divided into three parts: generation, transmission, and distribution. Distribution system (DS) is the first interface of the utility with the consumers and delivers electrical power to the end users. Due to a competitive environment and deregulation policies in the distribution sector, utilities are under continuous pressure to curtail operational cost by reducing power losses and improve reliability and other parameters to improve the overall performance. This has forced power system managers to make use of innovative practices to contribute in the evolution of the power system to a smart grid (SG). A smart grid is a modern, intelligent system that consists of sensors and monitoring mechanism, ICT, to give better performance and provide good support to consumers with economy.

Reconfiguration is altering of the topology of network by changing the ON/OFF (open/closed) status of the switches. It can be used to improve the operating conditions of the system [1–10].

In multiple attribute decision-making (MADM) [11–16], a number of alternatives (options) are considered along with some attributes (criteria) that are ranked and selected.

In this chapter, a subjective and objective integrated weightage (SOIW) method is used for finding the optimal configuration by considering some of the attributes from available alternatives for practical distribution systems, and results are compared with those of other MADM methods such as weighted sum method (WSM), weighted product method (WPM), analytic hierarchy process (AHP), and technique for order preference by similarity to ideal solution (TOPSIS) [17–22].

52.2 Decision-Making Methods

Decision-making is required for efficient and optimal use of all resources in any organization for every manager. In real-life situation, decision has to be made under complex conditions of several criteria/attributes, which are in contrast to each other, and this is a more challenging task. Hence, logical and systematic methods are required to be developed for assisting decision-makers by considering various attributes and their interrelations. The role of decision-maker is to identify appropriate selection attributes and get the most appropriate combination of attributes [23–28].

52.2.1 Multiple Attribute Decision-Making (MADM)

Each decision matrix or decision table in MADM methods consists of shortlisted alternatives, attributes, weightage, or importance of each attribute and measures of performance of alternatives. The decision matrix or table can be prepared as shown in Table 52.1. The decision table comprises alternatives, A_i (for $i = 1, 2, \dots, N$), attributes, B_j (for $j = 1, 2, \dots, M$), weightages of attributes, w_j (for $j = 1, 2, \dots, M$), and the measures of performance of alternatives, m_{ij} (for $i = 1, 2, \dots, N; j = 1, 2, \dots, M$) [11].

Table 52.1 Decision table or matrix by the MADM method [11]

Alternatives	Attributes					
	B1	B2	B3	–	–	BM
	(w1)	(w2)	(w3)	(–)	(–)	(wM)
A1	m11	m12	m13	–	–	m1 M
A2	m21	m22	m23	–	–	m2 M
A3	m31	m32	m33	–	–	m3 M
–	–	–	–	–	–	–
–	–	–	–	–	–	–
AN	mN1	mN2	mN3	–	–	Mnm

The role of the decision-maker is to select the best alternative from the given alternatives in the form of a decision table or matrix.

The most commonly used multi-criteria decision-making techniques are as follows:

- Weighted sum method (WSM).
- Weighted product method (WPM).
- Analytic hierarchy process (AHP).
- The technique for order preference by similarity to ideal solution (TOPSIS).
- Elimination and choice translating reality (ELECTRE).
- Preference ranking organization method for enrichment evaluations (PROMETHEE).

52.2.2 *Subjective and Objective Integrated Weightage MADM Method*

Subjective and objective integrated weightage multiple attribute decision-making was developed by Rao and Patel in 2010. Objective weights of importance of the attributes and the subjective preferences of the decision-maker are considered collectively to decide the integrated (aggregate) weights of the attributes. The proposed method helps the decision-maker to arrive at a decision based on either the objective weights of importance of the attributes or subjective preferences or taking into account both the objective weights and the subjective preferences.

The procedure of decision-making for solving distribution system problems using the subjective and objective integrated weightage (SOIW) method [15] is as follows:

Step 1: Identify and shortlist the alternatives on the basis of the identified criteria.

Step 2: Prepare a decision matrix or table after shortlisting the alternatives.

Step 3: Obtain normalized decision matrix m_{ij}^* by using Eq. (52.1).

$$m_{ij}^* = m_{ij} / \sum_{i=1}^N m_{ij} \quad (52.1)$$

where m_{ij}^* is the normalized value of m_{ij} and N .

Step 4: The weights of relative importance of the attributes can be decided by the decision-maker for the considered example based on three parameters, namely, objective weights of alternatives, subjective preferences of the attributes, or integration of both.

Objective weights of importance of the attributes: In this chapter, the statistical variance considered for determining the objective weights of importance of the attributes is calculated by using Eq. (52.2).

$$V_j = \left(\frac{1}{N}\right) \sum_{i=1}^N (m_{ij}^* - (m_{ij}^*)_{mean})^2 \tag{52.2}$$

where V_j is the statistical variance for the j th attribute and $(m_{ij}^*)_{mean}$ is the average value of m_{ij}^* .

The objective weight of the j th attribute (w_j^o) can be calculated by dividing the variance of j th attribute with the total value of the variances of “M” number of attributes. Thus, w_j^o can be found using Eq. (52.3).

$$w_j^o = \frac{V_j}{\sum_{j=1}^M V_j} \tag{52.3}$$

Subjective weights of importance of the attributes: The weights of relative importance of the attributes may be assigned based on the decision-maker’s preferences over the attributes for the considered application.

Integrated weights of importance of the attributes: The decision-maker can integrate both the objective and subjective weights of the attributes, using Eq. (52.4).

$$w_j^i = W^O w_j^o + W^S w_j^s \tag{52.4}$$

where w_j^i is the integrated weight of j th attribute, and W^S and W^O are the weightages given to the subjective and objective weights, respectively, and the values of W^S and W^O are between 0 and 1.

Step 5: The overall performance of an alternative is the weighted sum, which is called as preference index (P_i).

The preference index for each alternative indicates the score or merit of the alternative as compared to the other alternatives. The preference index (P_i) can be calculated using Eqs. (52.5), (52.6), and (52.7).

$$P_i^o = \sum_{j=1}^M w_j^o m_{ij}^{**} \tag{52.5}$$

$$P_i^s = \sum_{j=1}^M w_j^s m_{ij}^{**} \tag{52.6}$$

$$P_i^i = \sum_{j=1}^M w_j^i m_{ij}^{**} \tag{52.7}$$

where $m_{ij}^{**} = [m_{ij}^*b / (m_{ij}^*b)_{max}]$ for beneficial attributes and $[(m_{ij}^*nb)_{min} / m_{ij}^*nb]$ for nonbeneficial attributes. M_{ij}^*b and m_{ij}^*nb indicate the normalized values of the beneficial and nonbeneficial attributes, respectively.

Step 6: The alternative having the highest preference index value is the best choice for the considered decision-making problem.

Step 7: The final decision may be taken keeping in view practical considerations such as management constraints, social constraints, political constraints, etc. However, preference may be given for an alternative with a higher value of preference index.

52.3 Implementation and Results

Distribution System Case Study [18]

The application of the subjective and objective integrated weightage (SOIW) methodology proposed for a distribution system test network [18] based on an existing distribution network in an electricity distribution company is considered.

The test network consists of seven load centers having a concentrated load of 11 kV distribution network at each connection point, as well as 17 transformers and underground cables and overhead lines. In our case study, five different alternatives are to be evaluated by decision-makers. Attributes shortlisted for this case study are capital cost, annual energy losses, system security, supply availability, capacity constraints, and circuit length. All these attributes are required to be minimized as possible.

Application of Subjective and Objective Integrated Weightage (SOIW) Method for Decision-Making in a Smart Distribution System Case Study

The following procedure is followed for decision-making by using the subjective and objective integrated weightage MADM method:

Step 1: The selection attributes identified for selecting optimal switching combination are capital cost, annual energy losses, system security, supply availability, capacity constraints, and circuit length.

Step 2: The decision table or matrix is prepared and shown in Table 52.2.

Step 3: The data are normalized using Eq. (52.1), as the attributes have different units as shown in Table 52.3.

Step 4: The weights of relative importance of the attributes can be decided by the decision-maker by any one of the following:

- (a) The objective weights of the attributes are calculated using Eqs. (52.2) and (52.3) as follows:

$$w_{EL}^o = 0.000007, w_{SS}^o = 0.001784, w_{SA}^o = 0.014261, w_{CCONS}^o = 0.795030, \\ w_{CL}^o = 0.187471, w_{CCOST}^o = 0.001446.$$

- (b) The subjective weights considered in this work are taken as it is from reference [18]:

$$w_{EL}^s = 0.05, w_{SS}^s = 0.15, w_{SA}^s = 0.15, w_{CCONS}^s = 0.25, w_{CL}^s = 0.15, \\ w_{CCOST}^s = 0.25.$$

Table 52.2 Distribution system case [18]

Solution	Energy losses (MWh)	System security (% customers interrupted)	Supply availability (CML)	Capacity constraints (MWh)	Circuit length (km)	Capital cost (£'000)
1	14632.4	4.72	125.74	23.26	1.44	14,826
2	14584.3	4.69	120.22	109.8	1.59	14,813
3	14657.3	4.42	103.99	109.8	2.36	14,286
4	14647.8	4.93	141.74	23.26	0.68	15,703
5	14674.8	4.66	125.62	23.26	1.44	15,176

Table 52.3 Normalized data for DS case

Solution	Energy losses (MWh)	System security (% customers interrupted)	Supply availability (CML)	Capacity constraints (MWh)	Circuit length (km)	Capital cost (£'000)
1	0.1999	0.2015	0.2037	0.0804	0.1917	0.1982
2	0.1992	0.2003	0.1947	0.3794	0.2117	0.1980
3	0.2002	0.1887	0.1685	0.3794	0.3142	0.1910
4	0.2001	0.2105	0.2296	0.0804	0.0905	0.2099
5	0.2005	0.1990	0.2035	0.0804	0.1917	0.2029

Table 52.4 Integrated weights for different combinations

W^O	W^S	W^I_{EL}	W^I_{CI}	W^I_{SA}	W^I_{CC}	W^I_{CL}	W^I_{CCOST}
0	1	0.0500	0.1500	0.1500	0.2500	0.1500	0.2500
0.2	0.8	0.0400	0.1204	0.1229	0.3590	0.1575	0.2003
0.4	0.6	0.0300	0.0907	0.0957	0.4680	0.1650	0.1506
0.5	0.5	0.0250	0.0759	0.0821	0.5225	0.1687	0.1257
0.6	0.4	0.0200	0.0611	0.0686	0.5770	0.1725	0.1009
0.8	0.2	0.0100	0.0314	0.0414	0.6860	0.1800	0.0512
1	0	0.0000	0.0018	0.0143	0.7950	0.1875	0.0014

- (c) The integrated weights of importance of the attributes are computed using Eq. (52.4) and shown in Table 52.4 for different weightages given to the objective and subjective weights of the attributes.

Step 5: The preference index values are computed for different alternatives using Eqs. (52.5), (52.6), and (52.7) and listed in Table 52.5.

Step 6: The alternatives are ranked on the basis of preference index P_i values, and the highest P_i is rank 1, and so on (as shown in Table 52.6).

Step 7: A final decision may be taken by considering practical constraints (Fig. 52.1).

Results of WSM, WPM, and TOPSIS – MADM methods are shown in Tables 52.7, 52.8, and 52.9, respectively.

Table 52.5 Preference index for different weights

Solution	PiO	PiS	Pii						
			0.1	0.2,0.8	0.4,0.6	0.5,0.5	0.6,0.4	0.8,0.2	1,0
1	0.89842	0.87608	0.87608	0.88055	0.88502	0.88725	0.88949	0.89396	0.89842
2	0.26401	0.67933	0.67933	0.59627	0.51320	0.47167	0.43014	0.34708	0.26401
3	0.23993	0.69593	0.69593	0.60473	0.51353	0.46793	0.42233	0.33113	0.23993
4	0.99589	0.92176	0.92176	0.93658	0.95141	0.95882	0.96623	0.98106	0.99589
5	0.89842	0.87231	0.87231	0.87753	0.88276	0.88537	0.88798	0.89320	0.89842

Table 52.6 Rankings of solutions

Solution	W^O	W^S	Integrated weights W^I					$W^O = 1,$ $W^S = 0$
			$W^O = 0,$ $W^S = 1$	$W^O = 0.2,$ $W^S = 0.8$	$W^O = 0.4,$ $W^S = 0.6$	$W^O = 0.5,$ $W^S = 0.5$	$W^O = 0.6,$ $W^S = 0.4$	
1	3	2	2	2	2	2	2	3
2	4	5	5	4	5	4	4	4
3	5	4	4	4	4	5	5	5
4	1	1	1	1	1	1	1	1
5	2	3	3	3	3	3	3	2

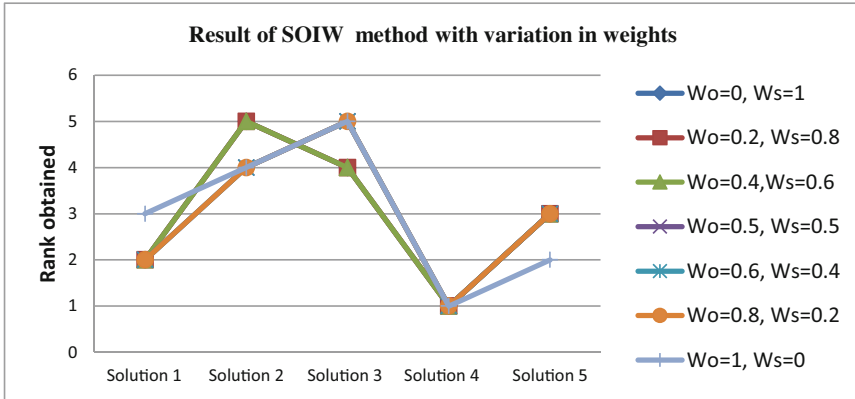


Fig. 52.1 Result of the SOIW method for distribution system case study

52.4 Conclusion

In this chapter, the subjective and objective integrated weightage (SOIW) MADM method is discussed in detail, and WSM, WPM, and TOPSIS methods are implemented for distribution system case study for decision-making.

For the distribution system case study, the attributes and weights considered are capital cost, annual energy losses, system security, supply availability, capacity constraints, and circuit length. The results obtained by all the MADM methods are compared, and solution number 4 has obtained the first rank in WSM, WPM, TOPSIS, and the proposed SOIW methods.

The subjective and objective integrated weightage method is simple and convenient and corresponds to the weighted sum method. The proposed SOIW method assists the decision-maker to conclude at a decision by considering (a) objective weights or (b) subjective weights or (c) integrated objective and subjective weights. The results obtained using the proposed SOIW method show a good correlation with other methods.

Table 52.7 Result of WSM for different weights

Solution	WO = 0, WS = 1		W = 0.2, WS = 0.8		W = 0.4, WS = 0.6		W = 0.5, WS = 0.5		W = 0.6, WS = 0.4		W = 0.8, WS = 0.2		WO = 1, WS = 0	
	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking
1	0.8761	2	0.8805	2	0.8850	2	0.8873	2	0.8895	2	0.8939	2	0.8984	3
2	0.6793	5	0.5963	5	0.5132	5	0.4717	4	0.4301	4	0.3471	4	0.2640	4
3	0.6959	4	0.6047	4	0.5135	4	0.4680	5	0.4223	5	0.3311	5	0.2399	5
4	0.9218	1	0.9366	1	0.9514	1	0.9588	1	0.9662	1	0.9811	1	0.9959	1
5	0.8723	3	0.8775	3	0.8828	3	0.8854	3	0.8880	3	0.8932	3	0.8984	2

Table 52.8 Result of WPM for different weights

Solution	WO = 0, WS = 1		W = 0.2, WS = 0.8		W = 0.4, WS = 0.6		W = 0.5, WS = 0.5		W = 0.6, WS = 0.4		W = 0.8, WS = 0.2		WO = 1, WS = 0	
	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking
1	0.8519	2	0.8547	2	0.8576	2	0.8591	2	0.8605	2	0.8634	2	0.8663	3
2	0.5740	4	0.4852	4	0.4102	4	0.3771	4	0.3467	4	0.2931	4	0.2478	4
3	0.5628	5	0.4708	5	0.3939	5	0.3603	5	0.3295	5	0.2756	5	0.2306	5
4	0.9170	1	0.9321	1	0.9475	1	0.9553	1	0.9632	1	0.9791	1	0.9953	1
5	0.8485	3	0.8521	3	0.8556	3	0.8574	3	0.8591	3	0.8627	3	0.8663	2

Table 52.9 Result of the TOPSIS method for different weights

Solution	WO = 0, WS = 1		W = 0.2, WS = 0.8		W = 0.4, WS = 0.6		W = 0.5, WS = 0.5		W = 0.6, WS = 0.4		W = 0.8, WS = 0.2		WO = 1, WS = 0	
	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking	Scores	Ranking
1	0.8028	2	0.8489	2	0.8763	2	0.8861	2	0.8941	2	0.9064	2	0.9152	3
2	0.2004	4	0.1533	4	0.1253	4	0.1153	4	0.1072	4	0.0948	4	0.0858	4
3	0.1368	5	0.0865	5	0.0543	5	0.0422	5	0.0322	5	0.0161	5	0.0044	5
4	0.8632	1	0.9135	1	0.9457	1	0.9578	1	0.9678	1	0.9839	1	0.9956	1
5	0.8014	3	0.8483	3	0.8760	3	0.8859	3	0.8940	3	0.9063	3	0.9152	2

References

1. Baran ME, Wu FF (1989) Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans Power Del* 4(2):1492–1498
2. Shirmohammadi D, Hong HW (1989) Reconfiguration of electric distribution networks for resistive line losses reduction. *IEEE Trans Power Syst* 4(2):1492–1498
3. Goswami SK, Basu SK (1992) A new algorithm for the reconfiguration of distribution feeders for loss minimization. *IEEE Trans Power Del* 7(3):1484–1491
4. Nara K, Shiose A, Kitagawa M, Ishihara T (1992) Implementation of genetic algorithm for distribution systems loss minimum re-configuration. *IEEE Trans Power Syst* 7(3):1044–1051
5. McDermott TE, Drezga I, Broadwater RP (1999) A heuristic nonlinear constructive method for distribution system reconfiguration. *IEEE Trans Power Syst* 14(2):478–483
6. Das D (2006) A fuzzy multiobjective approach for network reconfiguration of distribution systems. *IEEE Trans Power Delivery* 21(1):202–209
7. Siti MW, Nicolae DV, Jimoh AA, Ukil A (2007) Reconfiguration and load balancing in the LV and MV distribution networks for optimal performance. *IEEE Trans Power Delivery* 22(4):2534–2540
8. Srinivasa Rao R, Narasimham SVL, Ramalinga Raju M, Srinivasa Rao A (2011) Optimal network reconfiguration of large-scale distribution system using harmony search algorithm. *IEEE Trans Power Syst* 26(3):1080–1088
9. Amanulla B, Chakrabarti S, Singh SN (2012) Reconfiguration of power distribution systems considering reliability and power loss. *IEEE Trans Power Delivery* 27(2):918–926
10. Kavousi-Fard A, Niknam T (2014) Optimal distribution feeder reconfiguration for reliability improvement considering uncertainty. *IEEE Trans Power Delivery* 29(3):1344–1353
11. Rao RV (2007) *Decision making in the manufacturing environment*, pp 27–41. Springer-Verlag London Limited
12. Venkata Rao R, Patel BK (2009) Decision making in the manufacturing environment using an improved PROMETHEE method. *Int J Prod Res* 48:1–18. iFirst
13. Rao RV (2013) *Decision making in the manufacturing environment using graph theory and fuzzy multiple attribute decision making methods*, Springer series in advanced manufacturing, vol 2. Springer, London
14. Triantaphyllou E, Shu B, Nieto Sanchez S, Ray T (1998) Multi-criteria decision making: an operations research approach. In: Webster JG (ed) *Encyclopedia of electrical and electronics engineering*, vol 15. Wiley, New York, pp 175–186
15. Rao RV, Patel BK (2010) A subjective and objective integrated multiple attribute decision making method for material selection. *Mater Des* 31:4738–4747
16. IEEE Guide for Electric Power Distribution Reliability Indices, IEEE Std. 1366-2003, May 2004
17. Paterakis NG, Mazza A, Santos SF, Erdinç O, Chicco G, Bakirtzis A, Catalão JPS (2016) Multi-objective reconfiguration of radial distribution systems using reliability indices. *IEEE Trans Power Syst* 31:1048–1062
18. Tiefeng ZHANG, Guangquan ZHANG, Jun MA, Jie LU (2010) Power distribution system planning evaluation by a fuzzy multi-criteria group decision support system. *Int J Comput Intell Sys* 3(4):474–485
19. Kamble SG, Patil UV (2016) Performance improvement of distribution systems by using PROMETHEE – multiple attribute decision making method. International conference on communication and signal processing (ICCASP-2016)
20. Kamble SG, Patil UV, Vadirajacharya K Decision making in distribution system using different MADM methods. National conference on recent trends in electrical engineering at government college of engineering, Karad
21. Kamble SG, Vadirajacharya K, Patil UV (2017) Decision making in distribution systems using improved AHP-PROMETHEE method. IEEE international conference on Computing methodologies and communication (ICCMC 2017)

22. Kamble SG, Vadirajacharya K, Patil UV (2018) Comparison of multiple attribute decision-making methods—TOPSIS and PROMETHEE for distribution systems. International conference on communication and signal processing (ICCASP-2018) published in advances in intelligent systems and computing series https://doi.org/10.1007/978-981-13-1513-8_68
23. Ramanathan R, Ganesh LS (1990) A multi-objective programming approach to energy resource allocation problems. *Int J Energy Res* 17:105–119
24. Espie P, Ault GW, Burt GM, McDonald JR (2003) Multiple criteria decision making techniques applied to electricity distribution system planning. *IEE Proceed Generation Transm Distrib* 150 (5):527–535
25. Wong S, Bhattacharya K, Fuller JD (2009) Electric power distribution system design and planning in a deregulated environment. *IET Gener Transm Distrib* 3(12):1061–1078
26. Mazza A, Chicco G, Russo A (2014) Optimal multi-objective distribution system reconfiguration with multi criteria decision making-based solution ranking and enhanced genetic operators. *Int J Electr Power Energy Syst* 54:255–267
27. Pohekar SD, Ramachandran M (2004) Application of multi-criteria decision making to sustainable energy planning—a review. *Renew Sust Energy Rev* 8:365–381 369
28. Vitorino RM, Jorge HM, Neves LP (2013) Multi-objective optimization using NSGA-II for power distribution system reconfiguration” international transactions on electrical energy systems (2013) Wiley Online Library (wileyonlinelibrary.com). DOI: <https://doi.org/10.1002/etep.1819>

Chapter 53

Visual Importance Identification of Natural Images Using Location-Based Feature Selection Saliency Map



Malayil Abhayadev and T Santha

Abstract The proposed saliency map is called location-based feature selection saliency map (LBFSM). The research introduced a new method for identifying the image visual objects and region of unimportance. The saliency map uses Fourier transformation function for feature selection. The proposed method was applied over created natural images collected from different parts. The method's efficiency was calculated based on objective and subjective quality assessment matrices such as processing time, precision and recall values and receiver operator characteristic (ROC) values. The quality assessment study showed the proposed saliency method efficiency in finding the local and global features from the image. The performance of the state-of-the-art saliency calculation method was experimented on the same natural image dataset. Five different saliency maps and their performance were compared and evaluated based on subjective and objective measures. Nine hundred (CRIST900) natural images were experimented using MATLAB R2015a, and their quality assessment was done using the same software platform. This research gives a conclusion that the result of processing time, receiver operator characteristic (ROC) curve, precision, and recall values provide good performance compared to the state-of-the-art saliency map calculation methods.

Keywords LBFSM · ROC · Seam carving · Visual objects

Abbreviations

LBFSM	Location-based feature selection saliency map
ROC	Receiver operator characteristic
CRIST900	Content retargeting image resizing technique 900

M. Abhayadev (✉) · T. Santha
Department of Computer Science, Dr. G. R. Damodaran College of Science, Coimbatore, Tamil Nadu, India
e-mail: principal.cs@grd.edu.in

CSC	Continuous seam carving
r, g, b	Red, green, and blue
MOS	Mean opinion score
TPR	True-positive rate
FPR	False-positive rate
AUC	Area under the curve

53.1 Introduction

The objective of developing a computer vision system is to electrically duplicate the abilities of the human vision system. Saliency is the ability of a person to quickly focus on the most relevant parts of what he or she sees. What most people think important in an image is called image saliency. The resizing systems use a different variety of saliency map, and the overall performance of such systems is decreased by means of salient visible information preservation.

This saliency method is used in image processing for several years as presented by Chen Li-Qun et al. [1], Itti Laurent and Christof Koch [2], Itti Laurent et al. [3] and Ma Yu-Fei and Hong-Jiang Zhang [4]. Improved image resizing using seam carving and scaling better preserves the content of the image. The method considers frequency-tuned saliency map and distance to the centre of the image for each of the pixels. The saliency map is computed by gradient and visual saliency-based measure by Liu Zhi et al. [5]. Adaptive image retargeting using a saliency map is based on continuous seam carving (CSC), and it is a retargeting method. The method continuous seam carving assigns to each extracted seam a reserving ratio in accordance with the corresponding seam energy and reduces dimension by Liu Huiying et al. [6]. Region-based visual attention analysis with its application in image browsing on a small display was introduced by Liu Huiying et al.

The saliency measures should identify the important region in images. They construct an importance map, and this map ranks every pixel according to their visual importance by using gradient map, saliency map, and significance map. Humans automatically and effortlessly judge the importance in image regions and focus interest on important parts. Computationally detecting such salient image areas remains a vast goal, as it approves the preferential allocation of computational resources in subsequent image analysis and synthesis.

53.2 Related Work

The problem of visual saliency computation has proposed that the objective can be described as predicting, locating, and mining the salient visual information by simulating the corresponding mechanisms in the human vision system. The main

objective of saliency computation is to measure the importance of various visual subsets by popping out the targets and inhibiting the distracters. Saliency computation is carried out in three ways:

1. Location-based saliency category
2. Object-based saliency category
3. Learning-based saliency category.

Location-Based Category: The location-based saliency map is assumed to be assigned to specific locations. The saliency model separates high- and low-level frequencies based on the bottom-up feature selection method. The visual saliency values are assigned to various image object spatial locations like pixels and macro-blocks.

Object-Based Category: The object-based saliency map is directly assigned to objects. The importance of this map is that it concentrates on the object in an image. This method detects and analyses the salient object as a whole, and object saliency calculation is done on images by segmenting the image into two regions: Region of Importance and Region of Unimportance.

Learning-Based Category: The learning-based saliency map is calculated based on the learning-based category. This method is a top-down saliency calculation technology. The machine-learning feature selection improves the saliency map performance.

Tatler B et al.[7] introduced fixated and non-fixated location tracking technology. This research uses CRIST900 natural images, and these images are labelled into nine different groups. The proposed saliency map's efficiency was experimented on these natural image datasets. Many publically evaluable benchmark image datasets are used in feature selection research for the saliency map subjective evaluation process. Eye-tracking by Judd Tilke et al. [8] is the largest one in the community. It includes 1003 images collected from Flickr and LabelMe image databases. Bruce Neil and Tsotsos J [9] proposed a colour image benchmark. This benchmark contains 120 colour images from indoor and outdoor scenarios with a resolution of 681×511 .

53.3 Proposed Location-Based Visual Saliency Map (LBVSM)

The proposed location-based feature selection saliency map is used in this research for identifying the important object feature in the natural image dataset. Figure 53.1 shows the illustration of the proposed approach. A set of low-level and high-level visual features is extracted from the image dataset. The top-down saliency focuses on the high-level features like objects and human faces, while the bottom-up saliency focuses on low-level features. Both the methods identify the image feature using the Fourier transformation feature selection function; this function is fast and easy to implement. This map uses the feature transformation function to generate intensity,

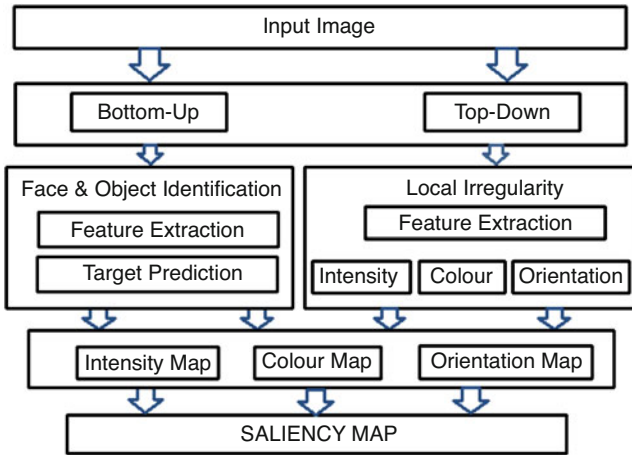


Fig. 53.1 The proposed visual saliency map feature selection framework

colour and orientation maps. The entire important salient object's boundary calculation is done in the target prediction stage. The proposed method identifies image object boundary most effectively. Finally, the three maps and the target predicted important objects combined generate the proposed saliency map on the natural images.

53.3.1 Visual Saliency Map Computation

The proposed model uses the location-based spatial domain saliency model, and the method follows a location-based visual saliency computation. High spatial frequencies are far more discriminatory than low spatial frequencies in contrast, and edge content and chromaticity are subject to the principle; the spatial saliency model domain computes visual saliency by detecting the irregularity in the input visual content. These irregularities are defined either locally, globally, or both. The local irregularity saliency model has three main modules of extraction for pre-attentive features, computation of multi-scale centre-surround contrasts and integration of contrast maps. Seven features are extracted from the input content, including intensity, red–green, and blue–yellow combinations and four orientations.

Image intensity feature I is calculated as

$$I = (r + g + b)/3 \quad (53.1)$$

where $r, g, b \in [0, 1]$.

The colour channels are represented as r, g, b (red, green and blue) channels of the input image. According to Hering's theory, some colours such as red and green as

well as blue and yellow cannot be interpreted at the same time, so these colour combinations are processed by the human vision system.

The normalized red, green, and blue channels can be computed as

$$\hat{r} = \begin{cases} r/I, & \text{if } I \geq 0.1 \\ 0, & \text{otherwise} \end{cases} \quad (53.2)$$

$$\hat{g} = \begin{cases} g/I, & \text{if } I \geq 0.1 \\ 0, & \text{otherwise} \end{cases} \quad (53.3)$$

$$\hat{b} = \begin{cases} b/I, & \text{if } I \geq 0.1 \\ 0, & \text{otherwise} \end{cases} \quad (53.4)$$

Four broadly tuned colours are created as

$$R = \max\left(0, \hat{r} - \frac{\hat{g} + \hat{b}}{2}\right), \quad (53.5)$$

$$G = \max\left(0, \hat{g} - \frac{\hat{r} + \hat{b}}{2}\right), \quad (53.6)$$

$$B = \max\left(0, \hat{b} - \frac{\hat{r} + \hat{g}}{2}\right), \quad (53.7)$$

$$Y = \max\left(0, \hat{r} + \hat{g} - 2(|\hat{r} - \hat{g}| + \hat{b})\right), \quad (53.8)$$

The red–green and blue–yellow opponencies can be calculated as

$$RG = R - G \quad (53.9)$$

and

$$BY = B - Y \quad (53.10)$$

The orientation features can be derived from the intensity channel by convolving it with Gabor filters in four directions.

Applying the convolution

$$O(\theta) = |I * G_0(\theta)| + \left|I * G_{\pi/2}(\theta)\right| \quad (53.11)$$

where $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ is the direction and

$$G_{\Psi}(x, y, \theta) = \exp\left(-\frac{x'^2 + Y^2 y'^2}{2\delta^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \Psi\right) \tag{53.12}$$

The Gabor filter has phase Ψ , aspect ratio Y , standard deviation δ , and wavelength λ . Coordinates (x', y') are transformed from (x, y) with respect to θ :

$$\begin{aligned} x' &= y \sin(\theta) + x \cos(\theta) \\ y' &= y \cos(\theta) - x \sin(\theta) \end{aligned}$$

The local irregularity saliency map identifies a feasible solution to detect saliency from an image centre region and saliency surroundings. Local contrast saliency can be used as an effective indication to quantize the difference between a region and its surroundings. The centre and the surrounding characteristics can be extracted using Difference of Gaussian function. The function takes different σ values to extract local irregularities from a different scale. Pre-attentive features such as *centre-surround* scales and contract maps are computed. The combined results are further normalized and summed up to generate three maps for intensity, colour and orientations. The local contrasts can well reveal the locations of salient targets from multiple scales. Finally, the three maps are combined to form the saliency map.

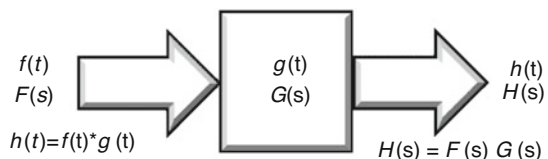
53.3.2 Transformation Function for Image Feature Selection

The transformation function takes the image as its input and produces an image as its output. This research uses Fourier transformation function for detecting high-level features from the image objects. A transformation function is an essential tool in the image processing research. In the Fourier spatial domain, each point in an image is represented as a particular frequency value. The Fourier transform is applied in many research areas such as image analysis, image compression and image resizing.

The natural image geometric characteristics are accessed with the help of Fourier transform. This domain breaks the image into sinusoidal modules and processes certain frequencies of the image by manipulating the local and global features of the images. Mathematical complexity of Fourier transform is less than that in Laplace transform and Z- transform. The Fourier transformation function is also called the spectrum of the input image (Fig. 53.2).

This is the basic principle behind any transformation function. The linear system is related to the convolution theorem.

Fig. 53.2 Linear system terminology



Here $f(t)$ = Input signal, $F(s)$ = Spectrum of input signal, $g(t)$ = Impulse response, $G(s)$ = Transfer function, $h(t)$ = Output signal, and $H(s)$ = Spectrum of output signal. The Fourier transform of one-dimensional function is

$$F(s) = \int_{-\infty}^{\infty} f(t)e^{-j2\pi st} dt \quad (53.13)$$

Fourier transform of Gaussian function

$$f(t) = e^{-\pi t^2} \quad (53.14)$$

Then Eq. 53.1 can be written as

$$F(s) = \int_{-\infty}^{\infty} e^{-\pi t^2} e^{-2\pi st} dt \quad (53.15)$$

The Fourier transform for two-dimensional image data, if $f(x, y)$ is an image and $f(u, v)$ is a spectrum, is

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{-j2\pi(ux+vy)} du dv \quad (53.16)$$

$$f(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux+vy)} dx dy \quad (53.17)$$

where u and v are two real frequency variables. The variable u corresponds to frequency along the x axis and v corresponds to frequency along the y axis.

The Fourier transform has several properties such as separability, similarity rotation, and projection. The separability property separates the two-dimensional image factors into one-dimensional component. The similarity property generates two-dimensional spectrums of images. The rotation property rotates the image through an angle θ and then the spectrum rotates the same amount. Another is the projection, which provides two-dimensional functions into one-dimensional function by x axis.

$$\text{If } f(x, y) = f_1(x)f_2(y) \quad (53.18)$$

Then,

$$= \int_{-\infty}^{\infty} f_1(x)e^{-j2\pi ux} dx \int_{-\infty}^{\infty} f_2(y)e^{-j2\pi vy} dy \quad (53.19)$$

The research proposed a transform-based location saliency map approach to estimate high- and low-level local features like intensity, red, green, and blue channels and angle of orientation in the image centre surroundings in the image. Colour features are gathered with the help of CIELAB colour ("Lab" colour) space.

This location-based saliency map is formulated by the feature selection transformation function and the input image is blurred with a difference of Gaussian filter. The normalized saliency map $S_n(x, y)$ is finally the combined saliency map, a combination of location-based feature selection with Fourier transform function and gradient information.

$$S_c = \alpha S_n + \beta S_g \quad (53.20)$$

where S_g is the gradient energy of the input natural image I and α and β are the scaling factors. The combined saliency result is the product of important image region and the image object edge features.

53.4 Implementation Stages for Proposed Visual Saliency Method

Step 1: Compute the location-based feature selection saliency map of the input image I to get S_C the final importance map.

Step 2: Compute the image I intensity value using Eq. (53.1), where $g, b \in [0, 1]$.

Step 3: Compute the image I normalized red, green and blue channels using eqs. (2), (3) and (4).

Step 4: Compute the orientation features from the intensity channels of image I by convolving it with Gabor filter in four directions using Eq. (53.12).

Step 5: Compute the local irregularly map by adding intensity, colour and orientation maps.

Step 6: Compute the high-level features of the image I using transformation function Eq. (53.16).

Step 7: Compute the final normalized saliency map $S_n(x, y)$,

$$S_C = \alpha S_n + \beta S_g$$

Here $S_n(x, y)$ is the combined top and bottom saliency map, α, β are the scaling factors. $S_g(x, y)$ is the gradient energy of the input image I .

53.5 Proposed Saliency Map Efficiency

The quality assessment evaluation is described as to find whether the saliency model can reflect the important visual content present in an image. The image dataset includes farmers, human faces, paddy field, nature, flowers, and so on. The evaluation metrics used in this research are receiver operating characteristic (ROC) as well

as recall and precision and mean opinion score (MOS). The receiver operating characteristic (ROC) takes to perform as a binary classifier and evaluates the performance of the saliency map. Based on the threshold value, the proposed saliency maps are binarized into forward and backward region, and this is done for calculating the true-positive, true-negative, false-positive, and false-negative values.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (53.21)$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (53.22)$$

The precision and recall values are used to detect the performance accuracy of saliency map objects. To calculate precision and recall values, an adaptive threshold is derived for each image to binarize the predicted saliency map.

The threshold value is fixed as

$$T_1 = \frac{2}{N} \sum_{n=1}^N S_n \quad (53.23)$$

By selecting the threshold value, a unique recall and precision can be calculated on all the experimented images. The performance F_β can be calculated as

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \quad (53.24)$$

β is a parameter to balance the influence of recall and precision matrices and $\beta^2 = 0.3$ in this research experiment. The proposed saliency map technique compares three existing methods: salient region detection and segmentation by Achanta Radhakrishna [10], context-aware saliency detection by Goferman S [11], and visual attention detection in video sequences using spatiotemporal cues by Zhai Yun and Mubarak Shah [12]. A model of saliency-based visual attention for rapid scene analysis is presented by Itti Laurent et al. [3].

53.6 Experimental Results

Figure 53.5 shows some worst result of the proposed saliency map. Some of the natural images with visual saliency important objects features are not considered by the proposed saliency map because some salient target must be locally important and exceptional, but this rare visual subset may not be always salient.

53.7 Comparison of Results with the State-of-the-Art Methods

The proposed saliency map has reliably outperformed all remaining state-of-the-art saliency calculation detection methods. The method gives higher precision and better recall rate for the created CRIST900 natural images. Figure 53.9 shows the Subjective Visual Comparisons of Different Saliency Maps. The Location-Based Feature Selection Saliency Map (LBFSM) fulfils the saliency map efficiency based on the following observations:

1. The location-based saliency calculation method separates the important objects from the surroundings and the gradient energy calculation in the natural images presented in Chap. 3 gives finite object boundary.
2. The method gives similar importance to objects and the region of unimportance in natural images and uniformly highlights entire important objects as well as the image background areas.
3. The saliency map's focus to the high contrast toward nearby regions and the low contrast to distant regions is less significant.
4. With the help of the proposed saliency map, the proposed hybrid content aware multi-operator works efficiently for natural image retargeting. The map is fast, easy to generate, and gives good results in the visual saliency computation.

Figures 53.3, 53.4, and 53.5 show the subjective quality assessment evaluation of the proposed saliency. The experimented natural images' object salient features are clearly extracted from the input images and shown in the output saliency map result. Visual comparison of saliency maps shows the qualitative and quantitative weaknesses of the state-of-the-art saliency map calculation approaches.

53.7.1 Processing Time Comparison

Figure 53.6 shows the proposed saliency map method and the other three saliency computation state-of-the-art methods processing time results.

53.7.2 Evaluation Metrics Receiver Operating Characteristics (ROC)

Receiver operating characteristic (ROC) is one form of an objective image quality assessment measurement. The Receiver operating characteristic curve is made by plotting the true-positive rate (TPR) against the false-positive rate (FPR) at various threshold settings. The true-positive rate is additionally called sensitivity, and the false-positive rate is additionally called the fall-out and calculated as $(1 - \text{specificity})$.

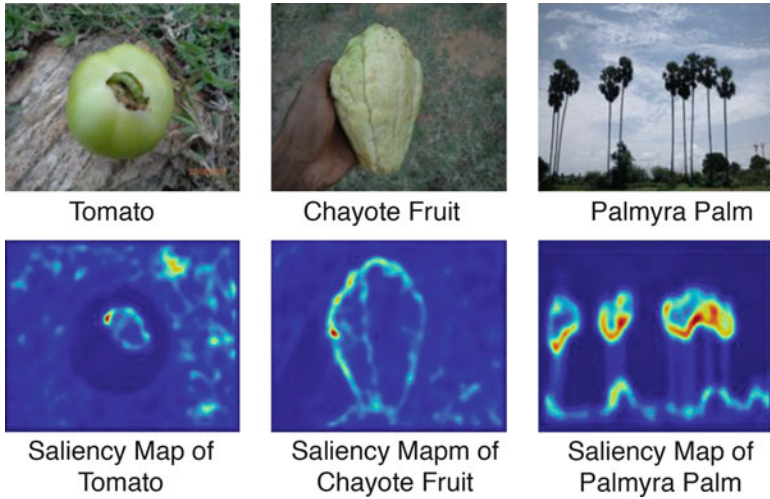


Fig. 53.3 Location-based feature selection saliency map experimental result 1. (top row left to right): Tomato image, chayote fruit image, and palmyra palm image. (bottom row left to right): saliency map of tomato image, saliency map of chayote fruit image, saliency map palmyra palm image

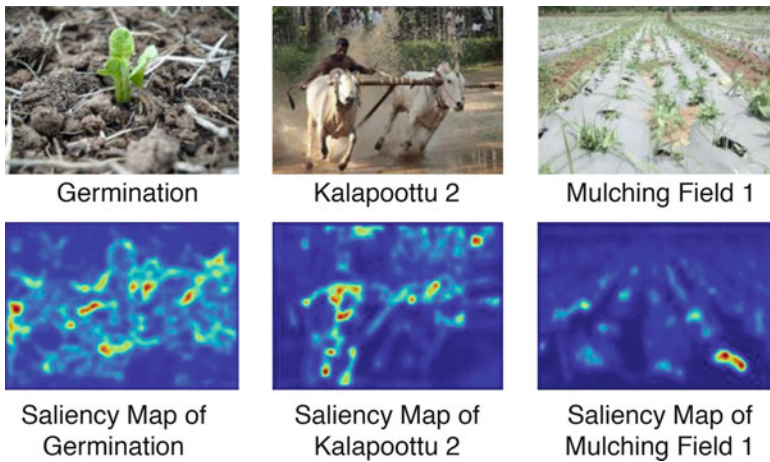


Fig. 53.4 Location-based feature selection saliency map experimental result 2. (top row – left to right): Germination image, kalapoottu 2 image, and mulching field 1 image. (bottom row-Left to Right): Saliency map of germination image, Saliency map of kalapoottu 2 image, and saliency map of mulching field 1 image

The ROC plot uses $1 - \text{specificity}$ on the x -axis and sensitivity on the y -axis. False-positive rate (FPR) is identical to $1 - \text{specificity}$, and the true-positive rate (TPR) is identical to sensitivity .

In Table 53.1, area under the ROC curve (AUC) score is an area under the curve calculated in the ROC space. The theoretical range of AUC score is between

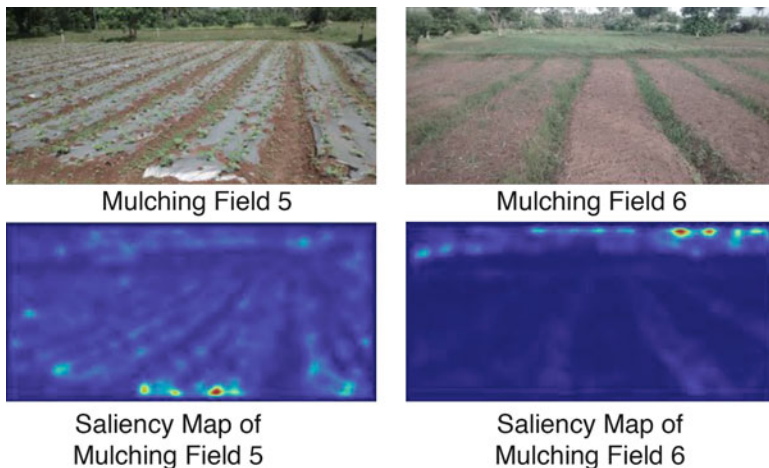


Fig. 53.5 Location-based feature selection saliency map experimental result 3. (top row – left to right): Mulching field 5 image, and Mulching field 6 image. (bottom row – left to right): Saliency map of mulching field 5 image, and saliency map of mulching field 6 image

Fig. 53.6 Graphical representation of saliency map processing time

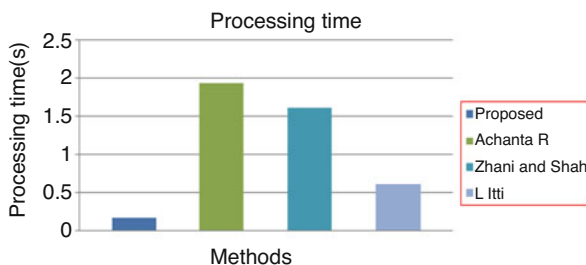


Table 53.1 Receiver operating characteristics (ROC) curve analysis

AUC	SE	95%	CI	Comment
0.86838	0.02801	0.81	0.92327	Good test

0 and 1, and the actual scores of meaningful classifiers are greater than 0.5. In Fig. 53.7, the AUC score is 0.8684; this shows that the performance level of ROC curves is good.

Figure 53.8 shows the resulting precision and recall curves of the proposed saliency map. The precision and recall curves clearly show that the proposed methods perform well. The saliency map is efficient for identifying visual importance saliency region in the natural images. $T_f \in [0, 255]$ is the threshold saliency map value and the maximum recall in which $T_f = 0$, all pixels are retained as positives and those pixels are considered as foreground. The maximum recall value is higher in the proposed method (Fig. 53.9).

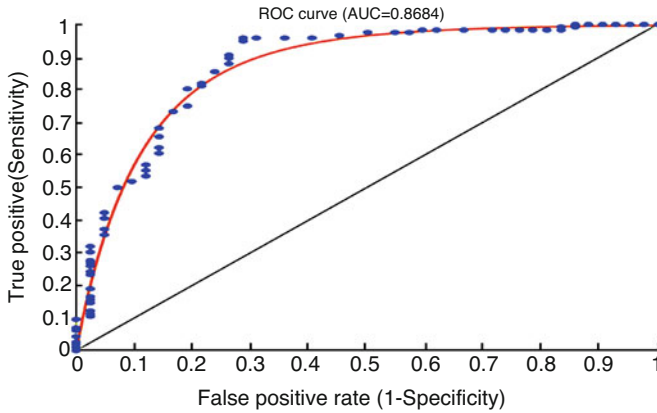


Fig. 53.7 Proposed saliency map receiver operating characteristics (ROC) graphs

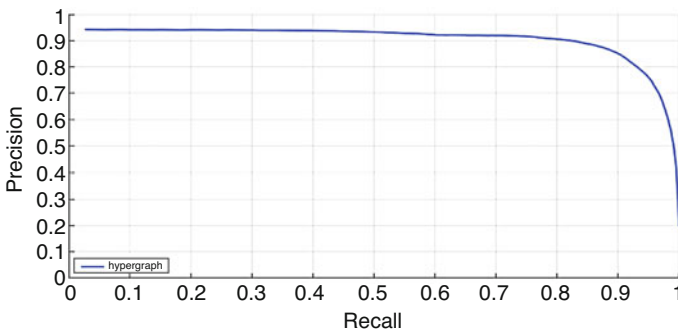


Fig. 53.8 Precision–recall curve

53.8 Conclusions

The research presents the location-based feature selection saliency map (LBFSM). The introduced saliency map is evaluated on the created natural image dataset and compares the proposed method efficiency with eight other state-of-the-art methods. Based on the processing time, precision, and recall values of the proposed saliency map is superior to the state-of-the-art methods.

53.9 Future Works

The future direction of our research process can be carried out by increasing the complexity of the image dataset. Future research can also be possible with videos.

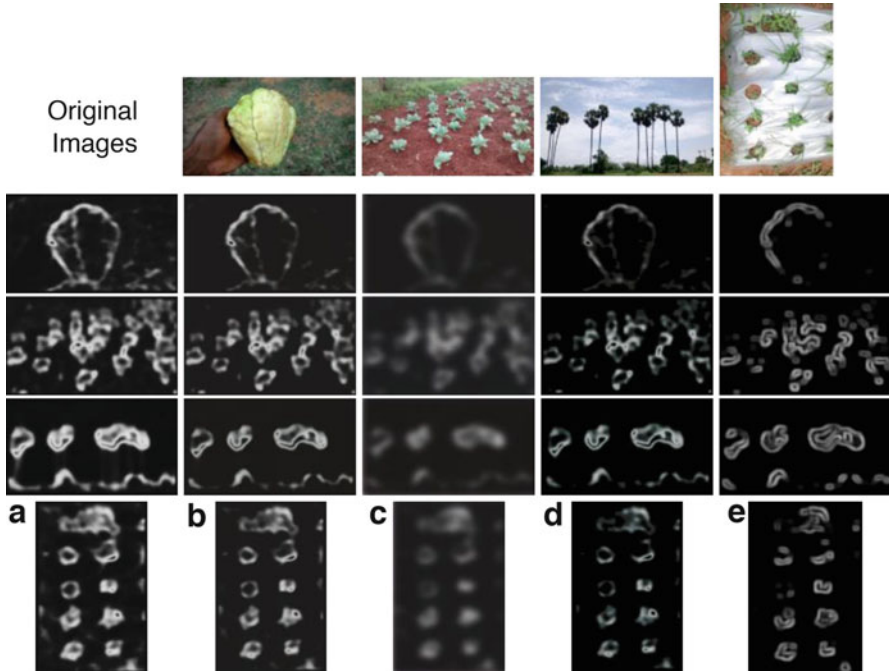


Fig. 53.9 The subjective evaluation of the proposed and the state-of-art saliency maps results. The first row (from left to right) shows original images (chayote fruit image, cauliflower image, palmyra palm image, mulching field 3 image). The first column shows (a) Proposed Saliency Map results; the second column shows (b) Saliency Map Result produced using Itti Laurent et al. [3]; the third column shows (c) Saliency Map produced using Goferman S [11]; the fourth column shows (d) Saliency Map produced using Achanta Radhakrishna et al. [10]; the fifth column shows (e) Saliency Map produced using Zhai Yun and Mubarak Shah [12].

References

1. Qun CL, Xie X, Fan X, Ma WY, Zhang H, Zhou HQ (2003) A visual attention model for adapting images on small displays. *Multimedia Systems* 9(4):353–364
2. Laurent I, Koch C (2001) Computational modelling of visual attention. *Nat Rev Neurosci* 2(3):194
3. Laurent I, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* 20(11):1254–1259
4. Fei MY, Zhang HJ (2003) Contrast-based image attention analysis by using fuzzy growing. In: *Proceedings of the 11th ACM international conference on multimedia*, pp 374–381
5. Zhi L, Yan H, Shen L, Ngan KN, Zhang Z (2010) Adaptive image retargeting using saliency-based continuous seam carving. *Opt Eng* 49(1):017002
6. Huiying L, Jiang S, Huang Q, Xu C, Gao W (2007) Region-based visual attention analysis with its application in image browsing on small displays. In: *Proceedings of the 15th ACM international conference on multimedia*, pp 305–308
7. Benjamin T, Baddeley RJ, Gilchrist ID (2005) Visual correlates of fixation selection: effects of scale and time. *Vis Res* 45(5):643–659

8. Tilke J, Ehinger K, Durand F, Torralba A (2009) Learning to predict where humans look. In: IEEE 12th international conference on computer vision, pp 2106–2113
9. Neil B, Tsotsos J (2006) Saliency based on information maximization. In: Advances in neural information processing systems, pp 155–162
10. Radhakrishna A, Estrada F, Wils P, Susstrunk S (2008) Salient region detection and segmentation. In: International conference on computer vision systems. Springer, Berlin/Heidelberg, pp 66–75
11. Goferman S, Zelnik L, Manor L, Tal A (2010) Context-aware saliency detection. In: Conference on computer vision and pattern recognition (CVPR), Vol. 1. No. 2
12. Yun Z, Shah M (2006) Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th ACM international conference on multimedia, pp 815–824

Chapter 54

Missing Values and Class Prediction Based on Mutual Information and Supervised Similarity



Nagalakshmi K. and S. Suriya

Abstract In recent times, the uses of data mining techniques have increased tremendously due to the increase in a large amount of data. Data mining techniques have been used for many research purposes. But mostly, they all face a single unique problem and that is the missing values of data. During research, large datasets are taken as processed for experimentation of algorithms, and if there is a missing value, these instances are either ignored or any default values are replaced during pre-processing of data. But this way is not correct. In this chapter, a novel prediction technique is proposed that can be used to predict the missing values of a given dataset or a dataset sample by calculating the mutual information, supervised similarity, and cosine similarity. The proposed approach calculated the missing values accurately, and this is experimented using a sample cancer dataset with missing gene values. The proposed prediction technique can also be used to predict class values of new instances of dataset. The experimentation shows that the predicted missing values and class labels coincide with the existing gene subsets and are said to be reliable and accurate.

Keywords Mining techniques · Class prediction · Supervised similarity · Mutual information · Missing value prediction

Abbreviations

ABC	Artificial bee colony
ACM	Association of computing machinery
GF	Gaussian function

Nagalakshmi K. (✉)
Sethu Institute of Technology, Virudhunagar, Tamil Nadu, India

S. Suriya
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_54

573

k-NN	K-nearest neighbor
MOABC	Multi-objective artificial bee colony
PDF	Probability density function
SVM	Support vector machine

54.1 Introduction

Data mining plays an important role in many applications due to the increase in large amounts of data these days. The amount of data that is generated every day from large institutions and organizations is really high, and many data mining techniques and algorithms [1] have been used to store, process, access, and handle these data. One such organization is the medical field that contains many hospitals and organizations, and the data here should be handled with care. Medical data are really sensitive and should be kept secure and this data should be made to provide useful information. But the number of attributes of any medical data is high, and so the dataset size is also big. Many studies have been performed in data mining in medical data, such as feature selection [2], feature classification [3], dimensionality reduction [4], etc. Apart from these, the concept or prediction also plays an important role in data mining in medical data.

Prediction [5] is the process of predicting something that is not already available in the data. To predict something, the existing data should be processed and analyzed. Then based on the obtained information, the prediction is made. In most cases of data mining research, the prediction is done to find the future value by using the current historical values. In other cases, the prediction concept can be used to predict unknown information by gathering information from the original dataset. But the major problem arises in data where there are missing values [6]. In current research methodologies, the records in the dataset with missing values are ignored during processing. For implementing data mining operations like classification and feature selection, the missing values should be ignored since they cannot be used. By ignoring such data, the obtained results cannot be compared with the original dataset.

This chapter provides a novel hybrid approach to predict missing values in a given dataset, especially medical dataset. This is because the medical records of a patient provide important information, and if the values of the medical test are missing, then it will be a major problem. Instead of ignoring the missing data record, they can be predicted using the other instance of the dataset. This way before applying data mining operations, the missing values can be predicted, and these values can be used. This is better than ignoring the data records completely. Apart from this, the proposed method also predicts the final output of the new instance of attributes or features provided. A novel hybrid method is used that makes use of mutual information and similarities between each of the features of the dataset to predict the missing data values.

The rest of the chapter is organized as follows: Section 54.2 describes the various existing methods and terminologies in prediction and feature extraction.

Section 54.3 provides details about the proposed method along with the various techniques and algorithms used. Section 54.4 shows the experimentation and results of the proposed method using a sample cancer dataset. Finally, Sect. 54.5 shows the overall conclusion and future work of the chapter.

54.2 Related Works

The proposed method in this chapter calculates the mutual information [7] about each feature of the dataset by using the probability densities of the features and the amount of entropy. The entropy and conditional entropies are used to calculate the mutual information [8]. Apart from this, the cosine similarity and supervised similarities are calculated for the features, and based on all the information, the predictions are made. In general, the existing methods make use of the mutual information and other similarity-based approaches for feature selection and classification. Some of the existing methods and related methods have been discussed below.

Hancer et al. [9] proposed multi-objective artificial bee colony (MOABC) for feature selection and classification.

Fuzzy-based mutual information is also proposed to evaluate the relevance between the feature subsets. Three different multi-objective feature selection approaches have been used here based on mutual information, fuzzy mutual information, and new fuzzy-based mutual information used here. The proposed method is experimented using a total of six common datasets, and comparison is made with existing single-objective ABC algorithms. The comparison shows that the proposed multi-objective approach performs better in terms of reducing the number of features.

Yu-Shuen Tsai et al. [10] provided a comparison between the feature selection strategies involved by calculating mutual information and fuzzy mutual information. The strategy used for calculating mutual information differs with different types of dataset and based on the needs. The mutual information is calculated differently for needs such as classification, feature selection, prediction, etc.; also, continuous dataset and hybrid datasets have different methods of mutual information calculation. The comparison indicates that the mutual information calculated using fuzzy provides more stable feature subset in feature selection than the normal mutual information.

WenhaoShu et al. [11] proposed a mutual information-based feature selection approach for set-valued dataset. In many data mining applications, the data generated from classification or analysis are set-valued. That is the feature subset of a given instance or an object will be denoted as set-valued, and this information will be useful in decision-making tasks. Many mutual information-based feature selection approaches and issues have been investigated for normal data, and no set-valued data issues are considered. A mutual information-based feature selection approach is proposed here for set-valued data.

Sehhati et al. [12] proposed a gene selection approach to select stable gene subsets from a given microarray dataset for the prediction of breast cancer occurrence. Here the prediction and selection are made by calculating a novel hybrid scoring method that is designed using linear combinations and is determined using mutual information and protein–protein interaction networks. Each of the genes in the cancer dataset is scored using the proposed scoring method using a hybrid forward–backward gene-set selection process, and the optimum gene subset is selected from the microarray dataset. Finally, the accuracy and stability of the gene selection are validated using the fivefold cross-validation.

Maji P. proposed [13] various f-information measures for effective selection of gene subsets from the given microarray datasets. Most of the feature selection approaches make use of the mutual information among the features to select the feature subsets. The mutual information also provides f-information that can be used for microarray data gene selection. The gene–gene redundancy is calculated here by measuring the divergence of joint distribution from the joint distribution when the genes are independent. The performances of various f-measures are compared with mutual information using classification accuracy from naïve Bayes, k-nearest neighbor, and support vector machine. Almost all f-information provided 90% or more accuracy.

54.3 Proposed Method

A novel prediction method is proposed in this chapter that is used to predict the missing values of already existing records of the dataset and the class values of new data instance. The proposed gene prediction approach predicts the missing gene values of the medical dataset [14] based on the relevance and redundancy between the attributes. The relevance between attributes [15] is calculated using the mutual information between the attributes. The redundancy [16] is measured using the similarity between attributes, and this is done using cosine similarity and a new supervised similarity measure. Finally, the missing values are predicted based on these three factors. Since three metrics are used for gene prediction, the accuracy of prediction will be high.

54.3.1 *Mutual Information*

First, the given dataset is split into C samples, where C is a used defined value to split the dataset into that many samples. The value of C can be taken as any value, and this will not affect the prediction. This is done in order to calculate different aspects of relevance and similarities between the attributes and finally combined together. In the dataset, a set of attributes will totally depend on other set of attributes in a functional dependency. Then the attributes have the same relevance with which they are related. This is termed

as mutual information between the attributes. The entropy values of the attributes are calculated, and based on this, the mutual information is obtained.

Let $C = \{c_1, 2, \dots, c_C\}$ be the number of samples in the dataset, and let $A = \{A_1, A_2, \dots, A_n\}$ be the number of attributes of gene values of the dataset, where n is the number of attributes. Let the number of output classes or the class labels of the dataset be D denoted as $D = \{D_1, 2, \dots, D_m\}$. The relevance of an attribute with respect to a given class label is defined as (D_j) where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$. The relevance of an attribute with respect to the class label determines how the attribute is related to that class label, that is, the amount of information gained using them together, which is mutual information. So the relevance value is calculated using mutual information as given in Eq. (54.1):

$$(D_j) = \text{Mutua}(A_i, D_i) \tag{54.1}$$

Here (A_i) determines the mutual information value between the attribute A_i and the class label D_j . The mutual information is calculated based on two measurements. They are the entropy and the conditional entropy [17]. The entropy of an attribute is determined as (A_i) and it defines the measure of uncertainty of the attribute within the dataset. The value of entropy is calculated by identifying the true probabilities of each possible values of the given attribute A_i . The true probabilities are calculated using a Gaussian function that takes as input the attribute value as well as the mean and standard deviation of the attribute within the given dataset or the given sample. The true portability of each attribute value a_{A_i} is given as (a) and the entropy is calculated as given in Eq. (54.2):

$$(A_i) = - \int (a) * \log p(a) da \tag{54.2}$$

After calculating the entropies of all the attributes or genes within all samples, the conditional entropy between each attribute $A = \{A_1, 2, \dots, A_n\}$ with each class values $D = \{D_1, D_2, \dots, D_m\}$ is calculated. The conditional entropy is determined as $(A_i|D_j)$, and it defines the amount of uncertainty left within the attribute A_i when knowing the class label as D_i , that is, the probability of the occurrence of both these values together in the samples is calculated using the probability density functions. The conditional probability density function of an attribute value $a \in A_i$ with respect to class variable $d \in D_j$ is determined as $p(a|d)$, and the joint probability density function of an attribute value $a \in A_i$ with respect to class variable $d \in D_j$ is determined as $p(a, d)$. The conditional probabilities and joint probabilities of all attributes with each class values are calculated, and the final conditional entropy is given as in Eq. (54.3):

$$(A_i, D_j) = - \int (a, d) * \log p(a|d) dadd \tag{54.3}$$

Finally, the mutual information between the attribute A_i and the class label D_j determined as (A_i, D_j) is calculated using entropy and conditional entropy as given in Eqs. (54.4) and (54.5):

$$Mu(A_i, D_j) = H(A_i) - H(A_i, D_j) \quad (54.4)$$

$$\text{MutualInfo}(A_i, D_j) = \iint p(a, d) \log p(a, d) p(a) * p(d) da dd \quad (54.5)$$

Here (a) and (d) determine the true probability density functions of each value of the attribute $a \in A_i$ and each values of the class label $d \in D_j$.

54.3.2 Supervised Similarity

The next step is the calculation of the redundancy between the attributes by using the similarity measure. The similarity is calculated only between two given attributes or gene values in the given dataset or dataset sample. The similarity or redundancy between two attributes (A_i, A_k) where $i = 1, 2, \dots, n$ and $k = 1, 2, \dots, n$ is also given as a measure of the mutual information between the two attributes as given in Eq. (54.6):

$$(A_i, A_k) = \text{Mutua}(A_i, A_k) \quad (54.6)$$

(A_i, A_k) is calculated using the same steps given in Eq. (54.2) up to Eq. (54.5). But as we can see here, the similarity incorporates only the redundancy or similarity between the attributes or genes, and the information about the class of the dataset or the sample is not incorporated. So, further, a supervised similarity measure is introduced here that calculates the redundancy or similarity based on the class labels also. The supervised similarity between two attributes is determined by (A_i, A_k) and is defined as in Eq. (54.7):

$$(A_i, A_k) = 11 + \lambda 2 \quad (54.7)$$

Here the value of λ is calculated based on the significance of an attribute with another attribute [18]. In this case, the significance of the attribute A_i with respect to the attribute A_k defined as (A_i) and the significance of the attribute A_k with respect to the attribute A_i defined as (A_k) are calculated. In general, the significance between two attributes is calculated as given in Eq. (54.8):

$$(Ak) = \{Ai, Ak\}(Dj) - RAi(Di) \quad (54.8)$$

The significance of an attribute is the measure of change in dependency of the attribute when the attribute is removed from the attribute set of the sample or dataset. The significance is higher with increase in dependency of the attribute and the significance is calculated with respect to the class labels. Based on this significance, the value of λ is given in Eqs. (54.9) and (54.10):

$$\lambda = (Ak) + \sigma Ak(Ai)2 \quad (54.9)$$

$$\lambda = \{Ai, Ak\}(Dj) - RAi(Di) - RAk(Di)2 \quad (54.10)$$

Now the similarity between two attributes can be calculated based on the significance of the attributes with each other combined with the class label. This way, the supervised similarity measure also takes into consideration the information about the dataset or sample class labels. By substituting the values of relevance with entropy and mutual information, the supervised similarity can be further extended as in Eq. (54.11)

$$\begin{aligned} \text{SuperSim}(Ai, Ak) = [1 + [\text{Sim}(Ai, Ak) - H(AiAk|Dj) - 1/2 \\ \{H(Ai) + H(Ai|Dj) + H(Ak) + H(Ak|Dj)\}]]2] - 1 \end{aligned} \quad (54.11)$$

From this, it is also evident that the supervised similarity also takes into account the similarities between the attributes that is defined before in Eq. (54.6) and also considers the entropies of each of the attributes with each class values.

54.3.3 Cosine Similarity

The mutual information calculates the relevancy between the attribute values, and the supervised similarity calculates the redundancy between the attributes of the given dataset or the given sample. But in order to predict a missing value, this information will not be enough. To predict missing values of any given instance or record of a dataset, the similar instance with this current missing instance should be calculated. The cosine similarity between two instances or gene subset can be calculated as given in Eq. (54.12):

$$(I1, I2) = I1.I2 / (|I1| * |I2|) \quad (54.12)$$

An instance or record in a dataset will contain many attribute values and each instance is represented as a vector. The similarity between two vectors or instance is

given in cosine similarity as given in Eq. (54.6). For example, consider that the dataset contains a total of 100 instances and each instance contains 10 attributes. Let the attribute value 5 in the instance 25 is missing, then the mutual information of attribute 5 with respect to each of the class labels should be first calculated. Then the supervised similarity between the attribute 5 and other attributes should also be calculated by considering all class labels of the dataset. Then finally, the cosine similarity between the instance 25 and each of the other instances should be calculated. Based on these three metrics, the missing value of attribute 5 in instance 25 can be predicted.

54.3.4 Class Prediction

Finally, the proposed method can also be used to predict the final class value given the sample new instance with the various attribute values. This is done by calculating the mutual information between the provided attribute values of the new instances with each class label of the dataset. Then cosine similarity is calculated between the new instance and each of the existing instances of the dataset. The new instance is then classified based on the calculated cosine similarity and mutual information. The class for which the mutual information is high and the class that is available with the existing instance to which the cosine similarity is high are predicted as the class of the new instance.

54.4 Implementation and Result

For the implementation purpose, the Wisconsin cancer dataset is taken as the input data. This dataset is chosen because it has been widely used for many data mining research in classification, feature selection, etc., but this dataset has many missing values. So in order to provide usefulness to the research of the upcoming researchers, this dataset has been chosen to predict the missing values. The sample values of the dataset and the various attributes and the class labels are shown in Fig. 54.1.

As seen from the dataset description from Fig. 54.1, the number of missing values in the dataset is from 16 instances, and a single gene or feature value is missing from them. It is denoted in the dataset using the symbol “?”. In most research, the authors ignore such instances or records and something they fill random values in them during pre-processing. The sample values of the dataset and the missing instances from the dataset are displayed in Figs. 54.2 and 54.3, respectively, for reference.

The dataset contains a total of nine features and two classes. The dataset is processed using the proposed prediction approach to calculate the mutual informa-


```

Number of Instances: 699 (as of 15 July 1992)

Number of Attributes: 10 plus the class attribute

Attribute Information: (class attribute has been moved to last column)

# Attribute                Domain
-----
1. Sample code number      id number
2. Clump Thickness         1 - 10
3. Uniformity of Cell Size 1 - 10
4. Uniformity of Cell Shape 1 - 10
5. Marginal Adhesion       1 - 10
6. Single Epithelial Cell Size 1 - 10
7. Bare Nuclei             1 - 10
8. Bland Chromatin         1 - 10
9. Normal Nucleoli         1 - 10
10. Mitoses                1 - 10
11. Class:                 (2 for benign, 4 for malignant)

Missing attribute values: 16

There are 16 instances in Groups 1 to 6 that contain a single missing
(i.e., unavailable) attribute value, now denoted by "?".

Class distribution:

Benign: 458 (65.5%)
Malignant: 241 (34.5%)
    
```

Fig. 54.1 Wisconsin cancer dataset description

```

1000025,5,1,1,1,2,1,3,1,1,2
1002945,5,4,4,5,7,10,3,2,1,2
1015425,3,1,1,1,2,2,3,1,1,2
1016277,6,8,8,1,3,4,3,7,1,2
1017023,4,1,1,3,2,1,3,1,1,2
1017122,8,10,10,8,7,10,9,7,1,4
1018099,1,1,1,1,2,10,3,1,1,2
1018561,2,1,2,1,2,1,3,1,1,2
1033078,2,1,1,1,2,1,1,1,5,2
1033078,4,2,1,1,2,1,2,1,1,2
1035283,1,1,1,1,1,1,3,1,1,2
1036172,2,1,1,1,2,1,2,1,1,2
1041801,5,3,3,3,2,3,4,4,1,4
1043999,1,1,1,1,2,3,3,1,1,2
1044572,8,7,5,10,7,9,5,5,4,4
1047630,7,4,6,4,6,1,4,3,1,4
    
```

Fig. 54.2 Sample dataset values

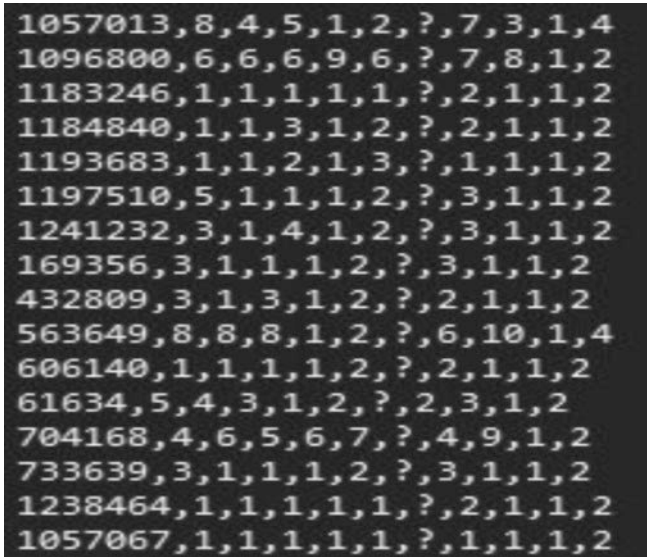


Fig. 54.3 Missing data instances in Wisconsin dataset

tion, supervised similarity measure, and the cosine similarity measures. The dataset is split into a total of two samples, with the first sample having a size of 444 instances and the second sample having a size of 239 instances, with a total of 683 instances from the 699 instances of the dataset. The missing values instances are ignored in these samples since they are to be predicted.

At first, the mean and standard deviations are calculated, and using this in the Gaussian function, the overall probability density functions are calculated. Then finally the entropy of each attribute is calculated. Similarly, the conditional entropies of all attributes with respect to the class labels are also calculated. By using the entropy and conditional entropy, the mutual information is calculated. All these values are displayed in Fig. 54.4.

In the next step, the supervised similarity measures are calculated as per the proposed method given in the previous chapter. The supervised similarity is calculated for each of the attributes of the Wisconsin cancer dataset with respect to the other attributes of the dataset. Similarly, the normal similarity or the redundancy measure is also calculated for the attributes with the other attributes. All these values are displayed in Fig. 54.5.

Finally, the prediction is made for the missing 16 instances of the given Wisconsin cancer dataset as shown in Fig. 54.6.

```

Conditional Entropy: Full Data
CondEnt(Attr1, Class 1) = 0.897
CondEnt(Attr1, Class 2) = 0.454
CondEnt(Attr1, Class 3) = 0.525
CondEnt(Attr1, Class 4) = 0.477
CondEnt(Attr1, Class 5) = 0.508
CondEnt(Attr1, Class 6) = 0.431
CondEnt(Attr1, Class 7) = 0.719
CondEnt(Attr1, Class 8) = 0.404
CondEnt(Attr1, Class 9) = 0.252
CondEnt(Attr1, Class 2) = 0.371
CondEnt(Attr1, Class 2) = 0.380
CondEnt(Attr1, Class 2) = 0.383
CondEnt(Attr1, Class 2) = 0.394
CondEnt(Attr1, Class 2) = 0.379
CondEnt(Attr1, Class 2) = 0.332
CondEnt(Attr1, Class 2) = 0.377
CondEnt(Attr1, Class 2) = 0.389
CondEnt(Attr1, Class 2) = 0.313

Mutual Information (Relevance): Full Data
MutInfo(Attr1, Class 1) = 147.876
MutInfo(Attr1, Class 2) = 144.296
MutInfo(Attr1, Class 3) = 145.918
MutInfo(Attr1, Class 4) = 152.655
MutInfo(Attr1, Class 5) = 176.685
MutInfo(Attr1, Class 6) = 127.472
MutInfo(Attr1, Class 7) = 162.472
MutInfo(Attr1, Class 8) = 146.077
MutInfo(Attr1, Class 9) = 211.453
MutInfo(Attr1, Class 2) = 148.402
MutInfo(Attr1, Class 2) = 144.370
MutInfo(Attr1, Class 2) = 146.061
MutInfo(Attr1, Class 2) = 152.738
MutInfo(Attr1, Class 2) = 176.814
MutInfo(Attr1, Class 2) = 127.571
MutInfo(Attr1, Class 2) = 162.814
MutInfo(Attr1, Class 2) = 146.092
MutInfo(Attr1, Class 2) = 211.392

Mutual Information (Relevance): Sample 1
MutInfo(Attr1, Class 1) = 142.682
MutInfo(Attr1, Class 2) = 83.295
MutInfo(Attr1, Class 3) = 155.532
MutInfo(Attr1, Class 4) = 159.840
MutInfo(Attr1, Class 5) = 142.334
MutInfo(Attr1, Class 6) = 159.846
MutInfo(Attr1, Class 7) = 143.816
MutInfo(Attr1, Class 8) = 134.726
MutInfo(Attr1, Class 9) = 84.828
MutInfo(Attr1, Class 2) = 143.128
MutInfo(Attr1, Class 2) = 83.369
MutInfo(Attr1, Class 2) = 155.675
MutInfo(Attr1, Class 2) = 159.923
MutInfo(Attr1, Class 2) = 142.463
MutInfo(Attr1, Class 2) = 159.945
MutInfo(Attr1, Class 2) = 143.358
MutInfo(Attr1, Class 2) = 134.740
MutInfo(Attr1, Class 2) = 84.767

Mutual Information (Relevance): Sample 2
MutInfo(Attr1, Class 1) = 43.385
MutInfo(Attr1, Class 2) = 53.935
MutInfo(Attr1, Class 3) = 51.838
MutInfo(Attr1, Class 4) = 59.765
MutInfo(Attr1, Class 5) = 46.174
MutInfo(Attr1, Class 6) = 46.077
MutInfo(Attr1, Class 7) = 36.714
MutInfo(Attr1, Class 8) = 53.157
MutInfo(Attr1, Class 9) = 37.301
MutInfo(Attr1, Class 2) = 43.911
MutInfo(Attr1, Class 2) = 54.610
MutInfo(Attr1, Class 2) = 59.848
MutInfo(Attr1, Class 2) = 46.383
MutInfo(Attr1, Class 2) = 46.196
MutInfo(Attr1, Class 2) = 39.056
MutInfo(Attr1, Class 2) = 53.172
MutInfo(Attr1, Class 2) = 37.241

CondEnt(Attr2, Attr1) = 0.137
CondEnt(Attr3, Attr1) = 0.144
CondEnt(Attr4, Attr1) = 0.131
CondEnt(Attr5, Attr1) = 0.134
CondEnt(Attr6, Attr1) = 0.120
CondEnt(Attr7, Attr1) = 0.157
CondEnt(Attr8, Attr1) = 0.123
CondEnt(Attr9, Attr1) = 0.080
CondEnt(Attr3, Attr2) = 0.144
CondEnt(Attr4, Attr2) = 0.131
CondEnt(Attr5, Attr2) = 0.134
CondEnt(Attr6, Attr2) = 0.120
CondEnt(Attr7, Attr2) = 0.157
CondEnt(Attr8, Attr2) = 0.123
CondEnt(Attr9, Attr2) = 0.080
CondEnt(Attr4, Attr3) = 0.131
CondEnt(Attr5, Attr3) = 0.134
CondEnt(Attr6, Attr3) = 0.120
CondEnt(Attr7, Attr3) = 0.157
CondEnt(Attr8, Attr3) = 0.123
CondEnt(Attr9, Attr3) = 0.080
CondEnt(Attr5, Attr4) = 0.134
CondEnt(Attr6, Attr4) = 0.120
CondEnt(Attr7, Attr4) = 0.157
CondEnt(Attr8, Attr4) = 0.123
CondEnt(Attr9, Attr4) = 0.080
CondEnt(Attr6, Attr5) = 0.120
CondEnt(Attr7, Attr5) = 0.157
CondEnt(Attr8, Attr5) = 0.123
CondEnt(Attr9, Attr5) = 0.080
CondEnt(Attr7, Attr6) = 0.157
CondEnt(Attr8, Attr6) = 0.123
CondEnt(Attr9, Attr6) = 0.080
CondEnt(Attr8, Attr7) = 0.123
CondEnt(Attr9, Attr7) = 0.080
CondEnt(Attr9, Attr8) = 0.080

Mean and Standard deviation: Full data
Attr1: Mean = 4.442, SD = 2.819
Attr2: Mean = 3.151, SD = 3.063
Attr3: Mean = 3.215, SD = 2.986
Attr4: Mean = 2.830, SD = 2.862
Attr5: Mean = 3.234, SD = 2.221
Attr6: Mean = 3.545, SD = 3.641
Attr7: Mean = 3.445, SD = 2.448
Attr8: Mean = 2.870, SD = 3.050
Attr9: Mean = 1.603, SD = 1.731

Average PDF: Full Data
Attr1: PDF = 0.898
Attr2: PDF = 0.893
Attr3: PDF = 0.894
Attr4: PDF = 0.183
Attr5: PDF = 0.136
Attr6: PDF = 0.875
Attr7: PDF = 0.116
Attr8: PDF = 0.895
Attr9: PDF = 0.280

Entropy: Full Data
Entropy(Attr1) = 148.773
Entropy(Attr2) = 144.759
Entropy(Attr3) = 146.444
Entropy(Attr4) = 153.132
Entropy(Attr5) = 177.193
Entropy(Attr6) = 127.992
Entropy(Attr7) = 163.190
Entropy(Attr8) = 146.481
Entropy(Attr9) = 211.705

Mean and Standard deviation: Sample 1
Attr1: Mean = 2.964, SD = 1.671
Attr2: Mean = 1.386, SD = 0.355
Attr3: Mean = 1.414, SD = 0.956
Attr4: Mean = 1.347, SD = 0.916
Attr5: Mean = 2.180, SD = 0.876
Attr6: Mean = 1.347, SD = 1.177
Attr7: Mean = 2.083, SD = 1.061
Attr8: Mean = 1.261, SD = 0.954
Attr9: Mean = 1.065, SD = 0.509

Average PDF: Sample 1
Attr1: PDF = 0.282
Attr2: PDF = 0.183
Attr3: PDF = 0.351
Attr4: PDF = 0.359
Attr5: PDF = 0.224
Attr6: PDF = 0.272
Attr7: PDF = 0.222
Attr8: PDF = 0.264
Attr9: PDF = 0.694

Entropy: Sample 1
Entropy(Attr1) = 143.499
Entropy(Attr2) = 83.749
Entropy(Attr3) = 156.858
Entropy(Attr4) = 151.317
Entropy(Attr5) = 142.842
Entropy(Attr6) = 151.277
Entropy(Attr7) = 143.735
Entropy(Attr8) = 135.129
Entropy(Attr9) = 85.080

Mean and Standard deviation: Sample 2
Attr1: Mean = 7.180, SD = 2.433
Attr2: Mean = 6.577, SD = 2.719
Attr3: Mean = 6.561, SD = 2.564
Attr4: Mean = 5.586, SD = 3.190
Attr5: Mean = 5.326, SD = 2.438
Attr6: Mean = 7.628, SD = 3.110
Attr7: Mean = 5.975, SD = 2.278
Attr8: Mean = 5.858, SD = 3.342
Attr9: Mean = 2.603, SD = 2.559

Average PDF: Sample 2
Attr1: PDF = 0.869
Attr2: PDF = 0.102
Attr3: PDF = 0.892
Attr4: PDF = 0.898
Attr5: PDF = 0.886
Attr6: PDF = 0.889
Attr7: PDF = 0.866
Attr8: PDF = 0.897
Attr9: PDF = 0.072

Entropy: Sample 2
Entropy(Attr1) = 44.282
Entropy(Attr2) = 54.390
Entropy(Attr3) = 52.364
Entropy(Attr4) = 51.242
Entropy(Attr5) = 46.682
Entropy(Attr6) = 151.277
Entropy(Attr7) = 39.433
Entropy(Attr8) = 53.561
Entropy(Attr9) = 37.554
    
```

Fig. 54.4 Entropy, conditional entropy, and mutual information calculations

Next, the class prediction is done for user-given inputs of a new gene subset or feature subset as given in Fig. 54.7.

54.5 Conclusion

In most data mining algorithms and research, the missing values of the input dataset are processed in many ways. Usually, the instances that contain the missing values are ignored, or in some cases, pre-processing is done to randomly fill the missing values. Sometimes, the users fill the missing values with a known value of another instance that is similar to that instance. This chapter proposes a novel missing value

Similarity (Redundancy): Full Data	Similarity (Supervised Similarity): Sample 1	Similarity (Supervised Similarity): Sample 2
Similarity(Attri 2, Attri 1) = -0.391	SuperSim(Attri 2, Attri 1) = 0.193	SuperSim(Attri 2, Attri 1) = 0.154
Similarity(Attri 3, Attri 1) = -0.391	SuperSim(Attri 3, Attri 1) = 0.424	SuperSim(Attri 3, Attri 1) = 0.307
Similarity(Attri 4, Attri 1) = -0.390	SuperSim(Attri 4, Attri 1) = 0.169	SuperSim(Attri 4, Attri 1) = 0.215
Similarity(Attri 5, Attri 1) = -0.391	SuperSim(Attri 5, Attri 1) = 0.005	SuperSim(Attri 5, Attri 1) = 0.005
Similarity(Attri 6, Attri 1) = -0.390	SuperSim(Attri 6, Attri 1) = 0.009	SuperSim(Attri 6, Attri 1) = 0.009
Similarity(Attri 7, Attri 1) = -0.392	SuperSim(Attri 7, Attri 1) = 0.018	SuperSim(Attri 7, Attri 1) = 0.021
Similarity(Attri 8, Attri 1) = -0.390	SuperSim(Attri 8, Attri 1) = 0.427	SuperSim(Attri 8, Attri 1) = 0.328
Similarity(Attri 9, Attri 1) = -0.388	SuperSim(Attri 9, Attri 1) = 0.001	SuperSim(Attri 9, Attri 1) = 0.001
Similarity(Attri 3, Attri 2) = -0.566	SuperSim(Attri 3, Attri 2) = 0.784	SuperSim(Attri 3, Attri 2) = 0.852
Similarity(Attri 4, Attri 2) = -0.566	SuperSim(Attri 4, Attri 2) = 0.061	SuperSim(Attri 4, Attri 2) = 0.063
Similarity(Attri 5, Attri 2) = -0.566	SuperSim(Attri 5, Attri 2) = 0.004	SuperSim(Attri 5, Attri 2) = 0.004
Similarity(Attri 6, Attri 2) = -0.565	SuperSim(Attri 6, Attri 2) = 0.013	SuperSim(Attri 6, Attri 2) = 0.013
Similarity(Attri 7, Attri 2) = -0.567	SuperSim(Attri 7, Attri 2) = 0.012	SuperSim(Attri 7, Attri 2) = 0.013
Similarity(Attri 8, Attri 2) = -0.565	SuperSim(Attri 8, Attri 2) = 0.761	SuperSim(Attri 8, Attri 2) = 0.790
Similarity(Attri 9, Attri 2) = -0.564	SuperSim(Attri 9, Attri 2) = 0.001	SuperSim(Attri 9, Attri 2) = 0.001
Similarity(Attri 4, Attri 3) = -0.492	SuperSim(Attri 4, Attri 3) = 0.093	SuperSim(Attri 4, Attri 3) = 0.099
Similarity(Attri 5, Attri 3) = -0.492	SuperSim(Attri 5, Attri 3) = 0.004	SuperSim(Attri 5, Attri 3) = 0.004
Similarity(Attri 6, Attri 3) = -0.492	SuperSim(Attri 6, Attri 3) = 0.011	SuperSim(Attri 6, Attri 3) = 0.011
Similarity(Attri 7, Attri 3) = -0.493	SuperSim(Attri 7, Attri 3) = 0.014	SuperSim(Attri 7, Attri 3) = 0.015
Similarity(Attri 8, Attri 3) = -0.492	SuperSim(Attri 8, Attri 3) = 0.940	SuperSim(Attri 8, Attri 3) = 0.902
Similarity(Attri 9, Attri 3) = -0.490	SuperSim(Attri 9, Attri 3) = 0.001	SuperSim(Attri 9, Attri 3) = 0.001
Similarity(Attri 5, Attri 4) = -0.201	SuperSim(Attri 5, Attri 4) = 0.007	SuperSim(Attri 5, Attri 4) = 0.007
Similarity(Attri 6, Attri 4) = -0.200	SuperSim(Attri 6, Attri 4) = 0.006	SuperSim(Attri 6, Attri 4) = 0.006
Similarity(Attri 7, Attri 4) = -0.202	SuperSim(Attri 7, Attri 4) = 0.039	SuperSim(Attri 7, Attri 4) = 0.043
Similarity(Attri 8, Attri 4) = -0.200	SuperSim(Attri 8, Attri 4) = 0.073	SuperSim(Attri 8, Attri 4) = 0.071
Similarity(Attri 9, Attri 4) = -0.198	SuperSim(Attri 9, Attri 4) = 0.001	SuperSim(Attri 9, Attri 4) = 0.001
Similarity(Attri 6, Attri 5) = 0.849	SuperSim(Attri 6, Attri 5) = 0.002	SuperSim(Attri 6, Attri 5) = 0.002
Similarity(Attri 7, Attri 5) = 0.848	SuperSim(Attri 7, Attri 5) = 0.019	SuperSim(Attri 7, Attri 5) = 0.018
Similarity(Attri 8, Attri 5) = 0.849	SuperSim(Attri 8, Attri 5) = 0.004	SuperSim(Attri 8, Attri 5) = 0.004
Similarity(Attri 9, Attri 5) = 0.851	SuperSim(Attri 9, Attri 5) = 0.003	SuperSim(Attri 9, Attri 5) = 0.003
Similarity(Attri 7, Attri 6) = -1.302	SuperSim(Attri 7, Attri 6) = 0.003	SuperSim(Attri 7, Attri 6) = 0.003
Similarity(Attri 8, Attri 6) = -1.300	SuperSim(Attri 8, Attri 6) = 0.012	SuperSim(Attri 8, Attri 6) = 0.012
Similarity(Attri 9, Attri 6) = -1.298	SuperSim(Attri 9, Attri 6) = 0.001	SuperSim(Attri 9, Attri 6) = 0.001
Similarity(Attri 8, Attri 7) = 0.238	SuperSim(Attri 8, Attri 7) = 0.014	SuperSim(Attri 8, Attri 7) = 0.013
Similarity(Attri 9, Attri 7) = 0.240	SuperSim(Attri 9, Attri 7) = 0.002	SuperSim(Attri 9, Attri 7) = 0.002
Similarity(Attri 9, Attri 8) = -0.488	SuperSim(Attri 9, Attri 8) = 0.001	SuperSim(Attri 9, Attri 8) = 0.001

Fig. 54.5 Similarity and supervised similarity

After Prediction:

8.0 - 4.0 - 5.0 - 1.0 - 2.0 - (4.0) - 7.0 - 3.0 - 1.0
6.0 - 6.0 - 6.0 - 9.0 - 6.0 - (6.0) - 7.0 - 8.0 - 1.0
1.0 - 1.0 - 1.0 - 1.0 - 1.0 - (1.0) - 2.0 - 1.0 - 1.0
1.0 - 1.0 - 3.0 - 1.0 - 2.0 - (1.0) - 2.0 - 1.0 - 1.0
1.0 - 1.0 - 2.0 - 1.0 - 3.0 - (1.0) - 1.0 - 1.0 - 1.0
5.0 - 1.0 - 1.0 - 1.0 - 1.0 - 2.0 - (1.0) - 3.0 - 1.0 - 1.0
3.0 - 1.0 - 4.0 - 1.0 - 2.0 - (1.0) - 3.0 - 1.0 - 1.0
3.0 - 1.0 - 1.0 - 1.0 - 2.0 - (1.0) - 3.0 - 1.0 - 1.0
3.0 - 1.0 - 3.0 - 1.0 - 2.0 - (1.0) - 2.0 - 1.0 - 1.0
8.0 - 8.0 - 8.0 - 1.0 - 2.0 - (8.0) - 6.0 - 10.0 - 1.0
1.0 - 1.0 - 1.0 - 1.0 - 2.0 - (1.0) - 2.0 - 1.0 - 1.0
5.0 - 4.0 - 3.0 - 1.0 - 2.0 - (4.0) - 2.0 - 3.0 - 1.0
4.0 - 6.0 - 5.0 - 6.0 - 7.0 - (6.0) - 4.0 - 9.0 - 1.0
3.0 - 1.0 - 1.0 - 1.0 - 2.0 - (1.0) - 3.0 - 1.0 - 1.0
1.0 - 1.0 - 1.0 - 1.0 - 1.0 - (1.0) - 2.0 - 1.0 - 1.0
1.0 - 1.0 - 1.0 - 1.0 - 1.0 - (1.0) - 1.0 - 1.0 - 1.0

Fig. 54.6 Final missing value prediction

```

Class Prediction for new input:

Enter a gene subset to predict class:
3
6
2
8
1
7
1
3
2

Classified as C2: 4.0: Malignant: Cancer

Do you want to give one more gene input? Yes or No?
Yes

Enter a gene subset to predict class:
1
2
1
3
4
1
2
1
3

Classified as C1: 2.0: Benign: No Cancer

Do you want to give one more gene input? Yes or No?

```

Fig. 54.7 Class prediction for new instances of genes

prediction technique that can be used to accurately predict the missing values of the given dataset based on mutual information, supervised similarity, and cosine similarity. Also, the proposed method can also be applied to predict the class of any given new instance based on the same information used to predict the missing values. The proposed method provides a new scope of research for prediction of missing values in real-time datasets for the use of research purpose. In future, this method can be employed to other types of datasets such as set-valued data and to datasets that contain more missing values with complicated gene subsets.

References

1. Padmapriya B, Velmurugan T (2014) A survey on breast Cancer analysis using data mining techniques. IEEE international conference on computational intelligence and research, December 2014, pp 1–4
2. Ang JC, Mirzal A, Haron H, Hamed HNA (2015) Supervised, unsupervised and semi-supervised feature selection: a review on gene selection. IEEE/ACM Trans Comput Biol Bioinform PP(99):1–1
3. Liu J-X, Xu Y, Zheng C-H, Kong H, Lai Z-H (2015) RPCA-based tumor classification using gene expression data. IEEE/ACM Trans Comput Biol Bioinform 12(4):964–970
4. Tang J, Zhou S (2016) A new approach for feature selection from microarray data based on mutual information. IEEE/ACM Trans Comput Biol Bioinform PP(99):1–1
5. Motai Y (2015) Kernel association for classification and prediction: a survey. IEEE Trans Neural Netw Learn Syst 26(2):208–223
6. Bose S, Das C, Dutta S, Chattopadhyay S (2012) A novel interpolation based missing value estimation method to predict missing values in microarray gene Expression data. IEEE international conference on communications, devices and intelligent systems, December 2012, pp 318–321
7. Pei Z, Zhou Y, Liu L, Wang L (2010) A mutual information and information entropy pair based feature selection method in text classification. IEEE international conference on computer application and system Modeling, October 2010, pp 258–261
8. Tian J, Wang Q, Bing Y, Dan Y (2013) A rough set algorithm for attribute reduction via mutual information and conditional entropy. IEEE 10th international conference on fuzzy systems and knowledge discovery, July 2013, pp 5667–571
9. Hance E, Xue B, Zhang M, Karaboga D (2015) A multi-objective artificial bee Colony approach to feature selection using fuzzy mutual information. IEEE congress on evolutionary computation, May 2015, pp 2420–2427
10. Tsai Y-S, Yang U-C, Chung I-F, Huang C-D (2013) A comparison of mutual and fuzzy-mutual information-based feature selection strategies. IEEE international conference on fuzzy systems, July 2013, pp 1–6
11. Shu W, Qian W (2014) Mutual information-based feature selection from set-valued data. IEEE 26th international conference on tools with artificial intelligence, November 2014, pp733–739
12. Sehhati M, Mehridehnavi A, Rabbani H, Pourhossein M (2015) Stable gene signature selection for prediction of breast Cancer recurrence using joint mutual information. IEEE/ACM Trans Comput Biol Bioinform 12(6):1440–1448
13. Maji P (2009) F-information measures for efficient selection of discriminative genes from microarray data. IEEE Trans Biomed Eng 56(4):1063–1069
14. Weitschek E, Felici G, Bertolazzi P (2013) Clinical data mining: problems, pitfalls and solutions. IEEE 24th international workshop on database and expert systems applications, August 2013, pp 90–94
15. Ebrahimpour M, Mahmoodian H, Ghayour R (2013) Maximum correlation minimum redundancy in weighted gene selection. IEEE international conference on electronics, computer and computation, November 2013, pp 44–47
16. Maji P, Das C (2012) Relevant and significant supervised gene clusters for microarray Cancer classification. IEEE Trans Nano Biosci 11(2):161–168
17. Dukkipati A, Pandey G, Ghoshdastidar D, Koley P, Sriram DMVS (2013) Generative maximum entropy learning for multiclass classification. IEEE 13th international conference on data mining, December 2013, pp 141–150
18. Alnemer LM, Al-Azzam O, Chitraranjan C, Denton AM, Bassi FM, Iqbal MJ, Kianian SF (2011) Multiple sources classification of gene position on chromosomes using statistical significance of individual classification results. IEEE 10th international conference on machine learning and applications and workshops, December 2011, pp 7–12

Chapter 55

Fake Product Review Detection and Removal Using Opinion Mining Through Machine Learning



Minu Susan Jacob, Selvi Rajendran, V. Michael Mario, Kavali Tejasri Sai, and D. Logesh

Abstract Machine learning is one of the growing trends in artificial intelligence and deep learning scenarios where the machine learns to acquire data from previous cases and implements the data for future prediction and analysis. The objective of this chapter is the detection and removal of fake reviews in online reviews. Majority of online buyers rely on product reviews before making purchase decision of their chosen brand; however, fake reviews pose a continuous threat to the integrity of the product, portals and the easy-to-find reviews on specific products. This chapter aims to develop a system to identify and remove fake reviews with the view of protecting the interests of customers, products and e-commerce portals. Thus, in this proposal, the primary goal is detecting unfair reviews on Amazon reviews through Sentiment Analysis using supervised learning techniques in an E-commerce environment. Sentiment classification techniques are used against a dataset (Amazon) of consumer reviews for smartphone products. Precisely, we use three different algorithms, logical regression algorithm, linear regression algorithm and neural networks (CNN and RNN models), of supervised machine learning technique to find similarities in the review dataset and group similar datasets together to explore unfair and fair positive and negative reviews, which involves screening, collaborative filtering, and removing with an optimal accuracy rate. The core focus or the highlight of this chapter is to explore an algorithm using deep learning that ensures optimal accuracy in the identification of fake reviews.

Keywords Sentimental analysis · Deep learning · Fake review · Neural networks (CNN and RNN)

M. S. Jacob (✉) · S. Rajendran · V. Michael Mario · K. T. Sai · D. Logesh
KCG College of Technology, Chennai, Tamil Nadu, India
e-mail: minu.cse@kcgcollege.com; selvi.cse@kcgcollege.com

Abbreviations

AUC	Area under the curve
CNN	Convolutional neural network
LSTM	Least short-term model
ML Algo	Machine learning algorithm
NLP	Natural language processing
RNN	Recurrent neural networks
WEV	Word embedding visualization
WITH ST	With stop words
WITHOUT ST	Without stop words

55.1 Introduction

Fake reviews are one of the major growing problems for those who own business. Fake reviews are similar to fake news, which is merely used by the developers to increase the advertisement review. Majority of the business transactions nowadays take place online. It is very easy to create a new account and post either positive or negative comments online for any business. A lot of costumers, when searching online for a product, will take a look at an online review or rating for a product. Fake reviews ultimately make the customer unhappy, as they are not getting the product they really wanted to get. Most of the research indicates that online reviews are too vague and may misguide consumers [1]. While we have tools and products to detect sources of fake news, the proposed approach considers the problem as an instance of text classification, using only the content of the article as the source of features. The focus is on NLP-related algorithms, thereby allowing us to explore in depth the performance of a variety of models on a particular task. Sentimental analysis is considered as the best technique to extract the subjective opinion of the text given. Sentiment analysis is a type of data mining that measures the inclination of people's opinions through natural language processing (NLP), computational linguistics and text analysis, which are used to extract and analyse subjective information from the Web – mostly social media and similar sources. The analysed data quantify the general public's sentiments or reactions towards certain products, people or ideas and reveal the contextual polarity of the information.

55.2 Related Work

This chapter examines the existing research and experimentation on the subject of opinion spam detection [1] and particularly in regard to fake reviews and reviewers [2]. It presents a variety of techniques, both supervised and unsupervised, in an

attempt to shed light on existing literature. Supervised methods rely mostly on content duplication and text categorization, with the latter gaining increased popularity after the release of the gold standard dataset. On the other hand, unsupervised techniques range from graph-based approaches to exploiting burstiness and deviation from mean ratings. The chapter is divided into two subsections according to the type of methods examined, which include spam review detection and spam reviewer detection. As stated in [3], the number of studies regarding review detection dominates the literature, while half as many focus on individual spammer detection [4]. Lastly, research on detecting spammer groups remains a rarity. By conducting several public opinion surveys, based on their results, it can be evaluated that people do read and get influenced by ratings and reviews of the products online [5]. A survey performed by a leading site has shown the following observations: More than 80% of the online customers look at the reviews available [6]. Fifty per cent base their purchase on the ratings of the products. Thirty per cent of the customers compare the ratings of similar products before making their decision.

Detection Techniques for Spam Reviews

The study examined a dataset of 5.8 million reviews and 2.14 million reviewers, crawled from Amazon [7]. Spam detection was performed as a classification problem consisting of two discreet classes: spam and non-spam. Detection of type 2 and 3 spam reviews was easily carried out by employing supervised learning through manually labelling spam and non-spam review instances as training data, owing to the obvious nature of these kinds of reviews, and building a logistic regression model based on review-centric, reviewer-centric and product-centric features. Due to the elusive content of type 1 reviews, however, labelling could not reliably take place; as such, 21 duplicate and near-duplicate reviews were examined, determined by measuring the Jaccard distances between review textual contents [10]. According to the research team, they are sure to contain type 1 spam reviews considering that the three distinct types of duplicate reviews include

- Duplicates from different users on the same product
- Duplicates from the same user on different products
- Duplicates from different users on different products

55.3 Methodology

55.3.1 Implementation Process

An end-to-end text classification pipeline is composed of the following components:

1. *Training Text*: It is the input text through which our supervised learning model is able to learn and predict the required class.
2. *Feature Selection*: A feature vector is a vector that contains information describing the characteristics of the input data (behavioural and linguistic model).

3. *Labels*: These are the predefined categories/classes that our model will predict.
4. *ML Algo*: It is the algorithm through which our model is able to deal with text classification (in our case: CNN, RNN, and logistic regression)
5. *Predictive Model*: A model that is trained on the historical dataset, which can perform label predictions.

This system mainly focuses on Amazon datasets of smartphone's product reviews. This chapter reviews data set pertaining to a smartphone brand of [Amazon.com](https://www.amazon.com), namely Redmi. The dataset used in our experiment is obtained from Amazon product data and divided based on a five-star rating: 1 star, 2 stars, 3 stars, 4 stars and 5 stars. The original dataset is not easy to model and usually not so clean. We have deleted some blank rows that cause confusion in the analysis process. The process uses neural network algorithms such as CNN and RNN to classify and group similarities/dissimilarities in the review data set, which classify both unfair and fair positive, neutral and negative reviews to get a comparison of results [8]. Similarities and dissimilarities are classified based on the sentiment expressed using various types of text patterns (with Stop words and without Stop words) including verbs, nouns and adjectives, etc. [9]. The algorithm helps filter and remove fake positive reviews and fake negative reviews with optimal accuracy as shown in Fig. 55.1.

1. *Analysing Our Data*

We use three types of dataset with various classes: Amazon product dataset (smartphone), hotel reviews dataset (TripAdvisor) and social media content as shown in Fig. 55.2.

2. *Data Processing*: In the dataset preparation step, we first perform tokenization. *Tokenization* is a process of extracting tokens (terms/words) from a corpus. Python's library Keras has an inbuilt model for tokenization, which can be used to obtain the tokens and their index in the corpus.

Processing Stage In this step, we first have to implement the following steps: 1. data mining classification, 2. text mining classification and 3. data-text mining classification. Now we apply each preceding step by more than one classification method.

Data Processing Involves the Following Steps:

1. Remove numeric and empty texts
2. Convert five classes into two classes (positive = 1 and negative = 0)
3. Remove punctuation from texts
4. Convert words to lower case
5. Remove Stop words
6. Stemming

The following is a list of brief contents of different parts:

- *Part-I*: In this part, I build a neural network with LSTM, and word embeddings were learned while fitting the neural network on the classification problem.

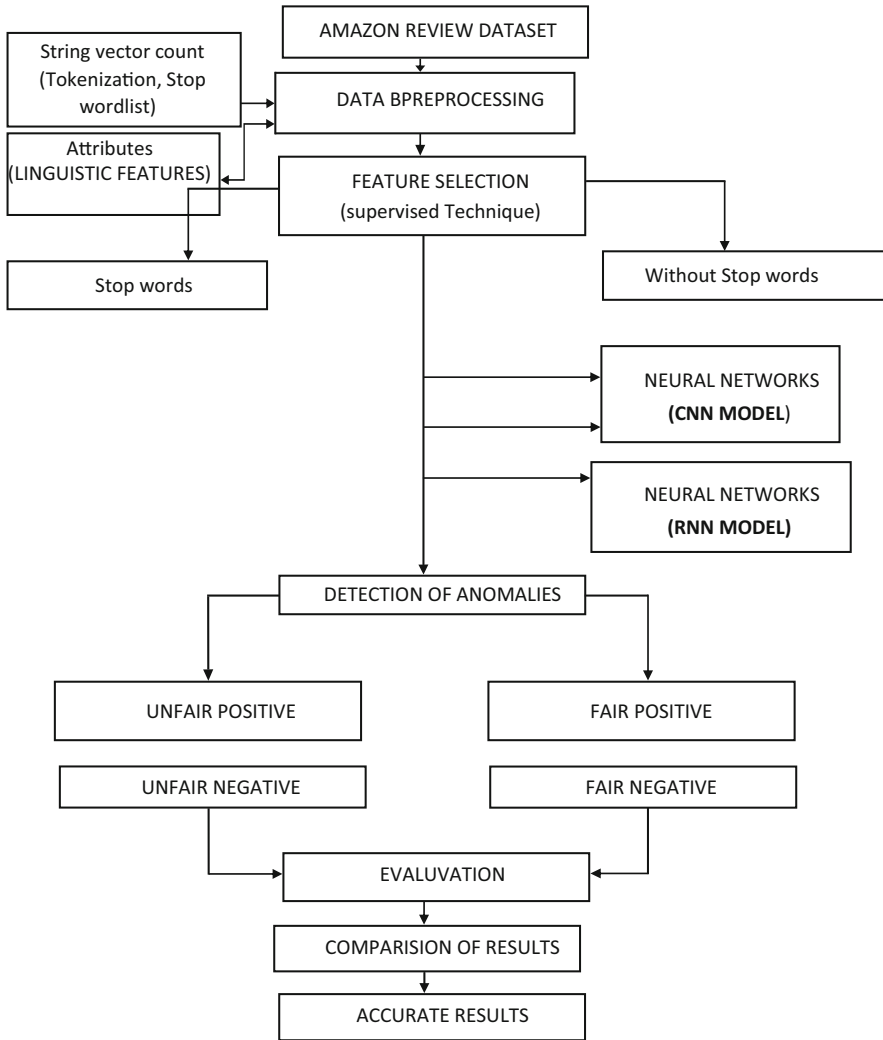


Fig. 55.1 Flow diagram

<u>DATASET</u>	<u>DATA SIZE</u>	<u>CLASSES</u>	<u>TRAIN/VALIDATION SAMPLES</u>
Dataset 2	18322	362	14658/3664
Dataset 1	3155	29	2524/631
Dataset 3	191	17	153/38

Fig. 55.2 Validation sample

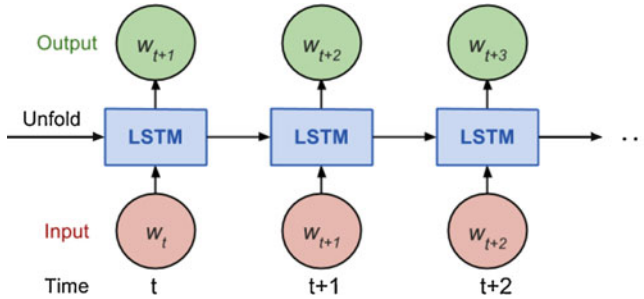


Fig. 55.3 LSTM

- *Part-2*: In this part, I add an extra 1D convolutional layer on top of the LSTM layer to reduce the training time.
- *Part-3*: In this part, I use the same network architecture as that in part 2 but use the pretrained glove 100-dimension word embeddings as the initial input.
- *Part 4*: In part 4, I use word2vec to learn word embeddings, as shown in Fig. 55.3.

Use pretrained glove word embeddings representation vectors into the LSTM model. This model in our case is essentially mapping a sequence of words to a class, and the workflow is shown below: *The mathematical equations for LSTM are as follows:*

input gate: $g u = \sigma(Wu * ht-1 + I u * xt)$, forget gate: $g f = \sigma(Wf * ht-1 + I f * xt)$
 output gate: $g o = \sigma(Wo * ht-1 + I o * xt)$, New memory cell: $g c = \tanh(Wc * ht-1 + I c * xt)$, Final memory cell: $mt = g f + g u g c$, Final hidden state: $ht = \tanh(g o mt-1)$.

Word Embedding Visualization

In this subsection, I want to visualize word embedding weights obtained from trained models. Word embeddings with 100 dimensions are first reduced to two dimensions using t-SNE. TensorFlow has an excellent tool to visualize the embeddings nicely, but here, I just want to visualize the word relationship.

55.4 Machine Learning Algorithms

This work deploys sentiment analysis following the supervised machine learning process to detect fake product review. This project has been adapted to review dataset pertaining to smart phone brand of [Amazon.com](https://www.amazon.com), namely, Redmi. The process uses logistic regression, convolutional neural networks and recurrent neural networks algorithms to classify and group similarities/dissimilarities in the review dataset, which classify both unfair and fair positive, neutral and negative reviews to get a comparison of result analysis [8]. Similarities and dissimilarities are classified

based on the sentiment expressed using various types of text patterns (with Stop words and without Stop words) including verbs, nouns and adjectives.

55.4.1 Convolutional Neural Network

The CNN model comprises *input layer, convolution layer, pooling layer and output layer*. We propose an improved CNN model considering the word order attributes. The input layer accepts the opinion sentences with a certain word order as input values. In the convolutional layer and the pooling layer, we maintain the word order of inputted sentences and adopt the word order persevering pooling method instead of the original pooling layer (LSMT MODEL).

Input Layer

We use word vectors to represent the word frequency of each word and take them as the training inputs of our model. We use the word2vec model to predict words that appear in the context by training a neural network language model to generate word vectors. The input layer consists of an $n \times m$ two-dimensional matrix, where n is the length of the sentence and m is the dimension of word vectors. The text representation process can be formulated as Eq. (1), where a represents the matrix, w represents the word vector of every word and v represents the value of every word vector. Ultimately, each opinion is represented by a two-dimensional word vector matrix.

Convolution Layer

The input layer transfers the word vector matrix A to the convolutional layer for convolutional operations. The padding of convolution is of two types: the same and valid. As is shown in Eq. 2, we perform the i th convolution in the l -layer K_l on matrix A , taking the ReLU function as activation function, the bias b_l as the valid padding of convolution and the matrix a_l as the feature map. The size of the convolution window is $h \times m$, where h is the width of the convolution kernel and m is the dimension of the word vector. The width of the convolution kernel (h) needs to be set and adjusted dynamically. As the convolution kernel continues to move down, the corresponding eigenvalues of the convolution kernel are generated. According to this convolution window, we will get a few of all '1' columns on the feature map.

The input value of the window is converted to an eigenvalue by the nonlinear transformation of the neural network. As the window moves down, the corresponding eigenvalues of the convolution kernel are generated and the eigenvectors corresponding to the convolution kernel are formed. We use the nonlinear transformation activation function called ReLU.

Word Order Persevering Pooling Layer

The word order persevering pooling layer reduces the number of feature parameters. The output of the order pooling layer is the maximum value of each feature map. The max pooling method can keep the location of the feature and the invariance value of the pooling operation. This feature affects the accuracy of text analysis, since the

texts exist with the word order characteristics. The position of each word in a sentence is a very important feature in the text analysis, so it is particularly important to preserve the word order of the sentences. Thus, the word order persevering the k-max pooling method is proposed here to replace the original max pooling method in the proposal. The order of the selected values corresponds to their original order in s . The word order-preserving k-max pooling method can discern more finely the number of times that the feature is highly activated in s than the max-pooling methods. What is more, the method can also distinguish the progression by which the high activations of the feature change across s . In this method, we can get the k highest feature values in the sequences.

Output Layer

We concat the obtained features from the pooling layer. It is a two-classification problem that distinguishes fake opinion from real opinion. The result of the concat function is then entered into the softmax function to assess the probability that the opinion is deceptive [11]. Finally, we use cross entropy as a model of the loss function to measure the difference between the predicted value and the true value in the OPCNN model as shown in Fig. 55.4.

55.4.2 Recurrent Neural Network (RNN)

A [recurrent neural network \(RNN\)](#) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit dynamic temporal behaviour for a time sequence. Using knowledge from an external embedding can enhance the precision of your RNN because it integrates new information (lexical and semantic) about the words, information that has been trained and distilled on a very large corpus of data. The pre-trained embedding we will be using is [GloVe](#). In this problem, while learning with a large number of layers, it becomes really hard for the network to learn and tune the parameters of the earlier layers. To address this problem, a new type of RNN called LSTM (**Long Short-Term Memory**) model has been developed.

LSTMs [12] have an additional state called ‘cell state’ through which the network makes adjustments in the information flow. The advantage of this state is that the model can remember or forget the learnings more selectively. To learn more about

	Real	Fake
Real	True negative reviews (TN)	True positive reviews (TP)
Fake	False positive reviews (FP)	False negative reviews (FN)

Fig. 55.4 The confusion matrix

```

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      539           70.1823 %
Incorrectly Classified Instances    229           29.8177 %
Kappa statistic                    0.3304
Mean absolute error                0.2988
Root mean squared error            0.5453
Relative absolute error            65.7327 %
Root relative squared error        114.3977 %
Total Number of Instances          768

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
Weighted Avg.  0.794    0.470    0.759     0.794    0.776     0.331    0.650    0.732    tested_negative
                0.530    0.206    0.580     0.530    0.554     0.331    0.650    0.469    tested_positive
Weighted Avg.  0.702    0.378    0.696     0.702    0.698     0.331    0.650    0.640

=== Confusion Matrix ===
   a  b  <-- classified as
397 103 | a = tested_negative
126 142 | b = tested_positive
    
```

Fig. 55.5 Weka tool

LSTMs, [here](#) is a great post. Let us architecture an LSTM model in our code. I have added a total of three layers in the model.

1. Input Layer: Takes the sequence of words as input
2. LSTM Layer: Computes the output using LSTM units. I have added 100 units in the layer, but this number can be fine-tuned later.
3. Dropout Layer: A regularization layer that randomly turns-off the activations of some neurons in the LSTM layer. It helps in preventing over-fitting.
4. Output Layer: Computes the probability of the best possible next word as output as shown in Fig. 55.5.

55.5 Detection Processes

After training, the next step is to predict the output of the model on the testing dataset and a confusion matrix generated, which classifies the review as positive or negative. We define as Fake the set of reviews that are found to be False (false positive or false negative) and defining as Real the set of reviews that are found to be True (true positive and true negative). The Fake and Real reviews are determined according to equations (a) through (d). The results involve the following attributes:

The detection process is accomplished by **With Stop words** and **Without Stop words of Fair and Unfair reviews**

- True Positive: Real positive reviews in the testing data, which are correctly classified by the model as Positive (P).
- False Positive: Fake positive reviews in the testing data, which are incorrectly classified by the model as Positive (P).
- True Negative: Real Negative Reviews in the testing data, which are correctly classified by the model as Negative (N).

- False Negative: Fake Negative Reviews in the testing data, which are incorrectly classified by the model as Negative (N)

as shown in Fig. 55.4.

The confusion matrix

$$\text{Fake Positive Reviews Rate} = \text{FP}/\text{FP} + \text{TN}$$

$$\text{Fake Negative Reviews Rate} = \text{FN}/\text{TP} + \text{FN}$$

$$\text{Real Positive Reviews Rate} = \text{TP}/\text{TP} + \text{FN}$$

$$\text{Real Negative Reviews Rate} = \text{TN}/\text{TN} + \text{FP}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

High Recall, Low Precision Indicates that most of the positive examples are correctly recognized (low FN), but there are a lot of false positives.

Low Recall, High Precision Indicates that we miss a lot of positive examples (high FN) but those we predict as positive are indeed positive (low FP).

F-score

It is the harmonic mean of the two values, which we have, that is, precision and recall.

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

It considers both the precision and recall of the procedure to compute the score. The higher the F-score, the better will be the predictive power of the classification procedure. A score of 1 means the classification procedure is perfect. The lowest possible F-score is 0 using the Weka tool.

Python Code Implementation Example: as shown in Fig. 55.5

Waikato Environment for Knowledge Analysis (Weka) is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand, and is free software licensed under the GNU General Public License. Weka supports several standard data mining tasks, more specifically, data pre-processing, clustering, classification, regression, visualization, and feature selection. All of Weka's techniques are predicated on the assumption that the data are available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other

attribute types are also supported). Aside from the main standalone Weka application, there is also a dedicated API for Java allowing the integration of Weka’s functionality, tools and algorithms directly in a Java project. The Weka API was used in this thesis for the implementation of the necessary classification tasks and evaluation.

55.6 Experimental Results

A gold-standard dataset for fake review detection is widely used for validating different models. However, since it is argued that the fake reviews written by the Amazon Mechanical Turk are not reliable, we attempted to create a dataset similar to the golden-standard dataset from the real-life dataset in (<http://liu.cs.uic.edu/download/data/>). This dataset is about the reviews from amazon.com, which is large and covers a very wide range of products. It is thus reasonable to consider it as a representative ecommerce site. The review dataset was crawled from amazon.com in June 2006; 5.8 million reviews, 2.14 reviewers and 6.7 million products are included in this dataset. We created the dataset based on Amazon dataset using the following steps. Using the above steps, we have collected 100 products where each product has 20 reviews. These 20 reviews are composed of eight fake reviews and 12 truthful reviews. The statistic information of the dataset is shown in Figs. 55.6 and 55.7.

When training the CNN model, we split the data into training, validation and testing sets, with an 80/10/10 split, and we again split sentences and conduct tokenization with NLTK. The two SVM-based models are trained according to the configurations.

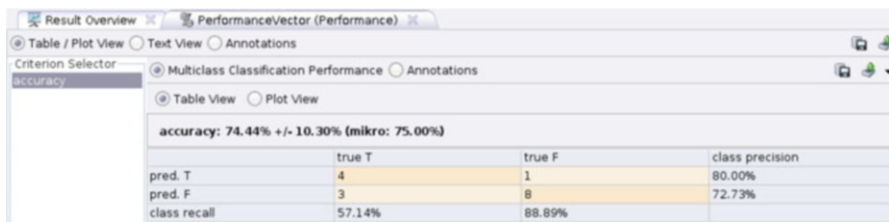


Fig. 55.6 Rapidminer results

Product	Number reviews	Number of deceptive reviews	Number of truthful reviews
100	2000	800	1200

Fig. 55.7 Sampling

When using the model, we set the widths of three convolutional filters as 1, 2 and 3. We learn 150-dimensional product-specific word embedding on each dataset; other parameters are initialized randomly from a uniform distribution. The KISS random search for hyper-parameters is adopted. To measure the overall classification performance, we use standard precision, recall and measure. Similarly, prediction is defined in Fig. 55.8.

After detection, the accuracy of the three various algorithms with their accurate precision and recall is assessed, and the number of positive reviews and negative Reviews is detected as shown in Fig. 55.9.

After successfully evaluating LSTM review text classification, the next section focuses on the main contribution of this thesis, the fake review detection model, which is based on the exploitation of spam indicators and author behaviour

Experimental method	Accuracy	Recall	F1-measure
CNN	80.33%	78.02%	82.79%
CNN-LSTM	74.20%	66.83%	68.76%
RNN-LSTM	68.27%	64.13%	65.85%
RNN	64.02%	68.22%	66.10%
Logical Regression	74.02%	77.01%	72.20%

Fig. 55.8 Comparativce study with various models

Train Sentiment	Precision	Recall	F-Measure	Accuracy
Positive (800 reviews)	0.862	0.862	0.862	86.25%
Negative (800 reviews)	0.866	0.865	0.864	86.5%
Combined (1600 reviews)	0.856	0.855	0.855	85.56%

Fig. 55.9 Precision recall measures

characteristics. The summary of the experiments, where CNN is the best algorithm by accuracy for all tests with Stop words and without Stop words. It can be inferred that CNN does not agree with other algorithms. CNN tends to be more accurate than other methods in comparison. The present study emphasizes that the accuracy of CNN tends to be higher when using the ‘without Stop words’ feature. However, the detection process of Fake Positive Reviews and Fake Negative Reviews offers less promising results than when using the ‘with Stop words’ feature as shown in Figs. 55.10 and 55.11.

In this paper, we proposed several methods to analyse a dataset of movie reviews and presented sentiment classification algorithms and supervised learning used in our work with Stop words and without Stop words methods. Our experimental approaches studied the accuracy of all sentiment classification algorithms and how to determine which algorithm is more accurate. Furthermore, we were able to detect

Features and Parameters	Fake Positive Reviews of SVM %	Fake Negative Reviews of SVM %	Precision of CNN%	Accuracy of CNN %	Accuracy of RNN %
Without stop words	18.8	17.7	81.4	81.75%	70.10%
with stop words	19.1	18.2	81.1	81.35%	69.90%

Fig. 55.10 Feature extraction

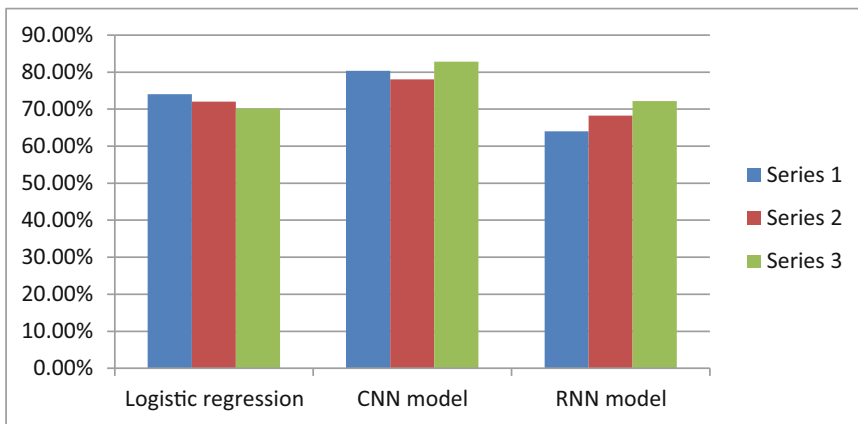


Fig. 55.11 Comparison with LR CNN RNN model

fake positive review and fake negative review through the detection processes as shown in our results. Four supervised learning algorithms to classify sentiment of our dataset have been compared in this chapter with Stop words and without Stop words. The first algorithm is logistic regression, the second algorithm is CNN, and the third algorithm is RNN. Through all of these algorithms, we have detected fake positive reviews and fake negative reviews. In this chapter, our experiments have shown the accuracy of results through sentiment classification algorithms, and we have found that the convolutional neural network algorithm in both cases, Stop words and without Stop words, is more accurate than other methods. Also, detection processes for fake positive reviews and fake negative reviews depend on the best and more accurate method used in this study. The main contributions of this study are summarized as follows: This study compares different sentiment classification algorithms in the Weka tool, which are used to classify movie reviews dataset into fake and real reviews. This study applies the sentiment classification algorithms without Stop words and with Stop words methods. We realized that with Stop words the method is more efficient not only in text categorization but also to detect fake reviews. This study performs several analysis and tests to find the best-supervised learning algorithm in terms of accuracy.

55.7 Conclusion

Since both implemented convolutional neural networks beat our baseline linear SVM in accuracy, the usage of CNNs for spam detection seems promising. Nevertheless, a character-level linear SVM application turned out to be surprisingly powerful, with lower complexity but almost comparable accuracy. For a final verdict about the CNN's applicability to spam classification, future research on larger spam datasets that are more representative of real-life email traffic is needed.

References

1. Anderson ET, Simester DI (2014) Reviews without a purchase: low ratings, loyal customers, and deception. *J Mark Res* 51(3):249–269
2. Ott M, Cardie C, Hancock J (2012) 2012. Estimating the prevalence of deception in online review communities. WWW
3. Wang Z (2010) Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews. *BE J Econ Anal Policy* 10(1):1–34
4. Heydari A, Ali Tavakoli M, Salim N, Heydari Z (2015) Detection of review spam: a survey. *Expert Syst Appl* 42(7):3634–3642
5. Sinha A, Arora N, Singh S, Cheema M, Nazir A (2018) Fake product review monitoring using opinion mining. *Int J Pure Appl Math* 119(12):13203–13209
6. Elmurungi EI, Gherbi A Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques. Received: 01-02-2018, Revised: 01-05-2018, Accepted: 11-05-2018

7. Li F, Huang M, Yang Y, Zhu X (2011) Learning to identify review spam. In: International joint conference on artificial intelligence, pp 2488–2493
8. Crawford M, Khoshgoftaar TM, Prusa JD, Richter AN, Najada HA (2015) Survey of review spam detection using machine identifying deceptive reviews based on labeled and unlabeled data. Ph.D. thesis, Wuhan University
9. Asadullah SM, Viraktamath S Classification of twitter spam based on profile and message model using svm
10. Seneviratne S, Seneviratne A, Kaafar MA, Mahanti A, Mohapatra P (2017) Spam mobile apps: characteristics, detection, and in the wild analysis. *ACM Trans Web* 11(1):129
11. Lupker SJ, Acha J, Davis CJ, Perea M (2012) An investigation of the role of grapheme units in word recognition. *J Exp Psychol Hum Percept Perform* 38(6):14911516
12. Olah C Understanding LSTM networks. Retrieved from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Chapter 56

Ask Less: Scale Market Research Without Annoying Your Customers



Venkatesh Umaashankar and S. Girish Shanmugam

Abstract Market research is generally performed by surveying a representative sample of customers with questions that include contexts such as psychographics, demographics, attitude, and product preferences. Survey responses are used to segment the customers into various groups that are useful for targeted marketing and communication. Reducing the number of questions asked to the customer has utility for businesses to scale the market research to a large number of customers. In this work, we model this task using Bayesian networks. We demonstrate the effectiveness of our approach using an example market segmentation of broadband customers.

Keywords Market research · Market segmentation · Bayesian networks · Graphical models · Dimensionality reduction · Survey

Abbreviations

AIC	Akaike information criterion
B2C	Business to customer
BIC	Bayesian information criterion
BN	Bayesian network
DAG	Directed acyclic graph
ISP	Internet service provider

V. Umaashankar (✉)
Ericsson Research, Chennai, India
e-mail: venkatesh.u@ericsson.com

S. Girish Shanmugam
Uppsala University, Uppsala, Sweden

56.1 Introduction

A key technique for developing successful business strategies in business-to-customer (B2C) companies is to develop a good understanding of the market and the customer behavior. Market research and segmentation play an important role in framing business and marketing strategies, which help organizations to improve the efficiency of their marketing and conversion. Market segmentation could be defined as the process of breaking down the market for a particular product or service into segments of customers, which differ in terms of their response to marketing strategies [1].

Market segmentation comprises two major steps: (1) consumer survey—a survey questionnaire considering various dimensions, such as psychographics, demographics, attitude, product usage, and preferences, is meticulously designed. Psychographic questions are useful in understanding the preferences and behavior of customers [2]. The carefully planned survey is then rolled out to a representative sample of customers. (2) Segment generation—the survey responses are analyzed to create a segmentation model. The segmentation model could be based on a rule, an algorithm, or factor analysis. This model can be abstractly defined as a function that could be used to assign a segment to every surveyed customer. The process of market segmentation is discussed in detail in [3].

$$SegmentationModel(responses) = segment \tag{56.1}$$

The market segments are summarized by profiles and are given descriptive names. Consider the example of market segmentation of movie consumers [4] shown in Fig. 56.1. This market research was done for a studio to understand the

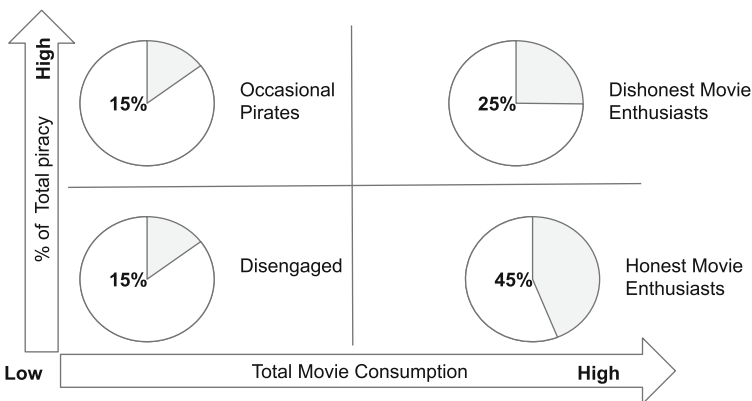


Fig. 56.1 Movie consumers’ market segmentation

Table 56.1 Segment descriptions

Occasional pirates	Disengaged	Dishonest enthusiasts	Honest enthusiasts
Higher conversion potential	Moderate conversion potential	Low conversion	High potential
Pirating family/ action and low risk to pirate	Low risk to pirate	Action preferred genre	Across all genres
Females aged 25–34 years with young family	45+ years family	Younger male students	35–44 with family

level of piracy among movie consumers. They report four clear market segments classified based on consumption level and tendency to consume pirated material.

The heterogeneity among the segments is emphasized in the descriptions shown in Table 56.1. Segment descriptions help to build an intuition about the nature and behavior of each segment. Market segmentation for a product or service is usually executed by expert market research companies.¹ The key outcomes of market research are segmentation model, target segments, presentations, and workshops to spread the awareness within the organization. Market segmentation has been battle-tested in many consumer-facing businesses, and it clearly helps to build the intuition about the big picture. Still, it is an open challenge to scale market research to millions of customers. It is not practical to ask a long list of questions to each and every customer that would not only be time-consuming but also be annoying the customers.

Factor analysis is a well-known method for estimating the latent traits from question-level survey data and to reduce the number of questions [5]. However, it has also been the subject of no small amount of criticism among market researchers [6]. The major problem with factor analysis is that we lose the diversity in the collected information, and we will have only minimal information. A factor analysis carried out on one-half of the data might give different results from those obtained from the other half, thus making the reliability of results questionable. Yet another limitation is that it is unable to give a unique solution or result. An exercise in factor analysis involving a large number of variables, say 50, is much bothersome, costly, and time-consuming [7]. Due to these limitations, we avoided factor analysis and decided to go for a much simpler alternative.

A Bayesian framework that systematically addresses the challenges faced when the future value of customers is estimated based on survey data have been proposed in [8]. A method for building effective Bayesian network (BN) models for medical decision support from complex, unstructured, and incomplete patient questionnaires and interviews was developed in [9]. It extends to challenging the decision scientists to reason about building models based on what information is really required for inference.

¹Ipsos and TNS are well-known market research experts in the industry.

The closest to our work is [10], where Bayesian network modeling has been used instead of applying factor analysis technique to determine key factors from a survey questionnaire to find the most accurate representation of the complex system and identify key variables for understanding the subsequent effects of blast exposure based on an online survey. To the best of our knowledge, there has not been any other work exploring the use of the Bayesian network for scaling market research or to reduce the number of questions in a market research survey.

In this work, we propose a novel way to use Bayesian networks to reduce the number of questions that a customer needs to be asked. In addition to that, we demonstrate the effectiveness of our approach by evaluating the segment assigned by the Bayesian network model when fewer questions are asked in the survey. Finally, we summarize the advantages of our approach and discuss our conclusions.

56.2 Proposed Approach

Inspired by the success of using Bayesian networks to understand and analyze survey data [11], we propose a Bayesian network-based approach for reducing the number of questions in a market research survey. The outline of our approach is shown in Fig. 56.2. Our approach consists of two phases: (1) Preparatory phase and (2) scaling phase.

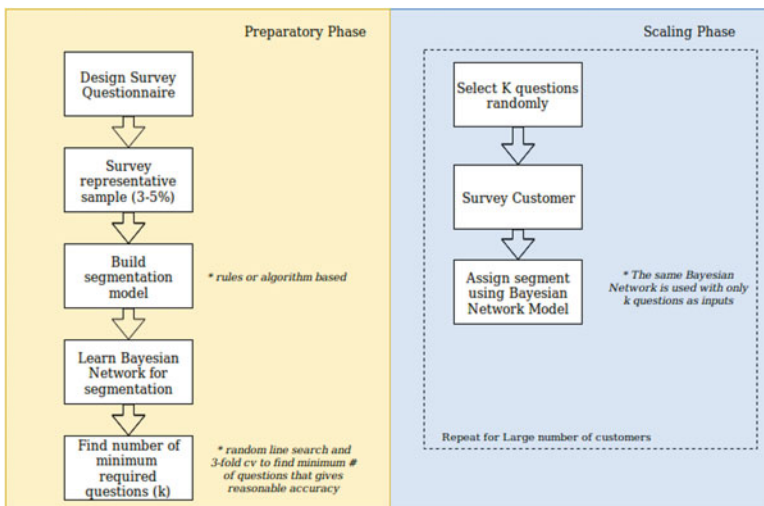


Fig. 56.2 Ask less – approach outline

56.2.1 Preparatory Phase

In the preparatory phase, the survey questionnaire is designed and the survey is rolled out to a representative sample of customers, which is typically 2–5% of the total customer base. Customers selected for these phases are usually sampled in stratified fashion across various regions of value that they add to the business. The survey responses are analyzed, and a segmentation model is built to divide the customers into different segments. A segmentation model could be defined as a function that takes survey responses as input and provides customer segment as output as shown in Eq. 56.1. Note that the questions have to be carefully designed, keeping in mind what type of segments the business would benefit from. Also, note that it might be the case that all the survey responses are similar, and it might not be possible to differentiate customers based on the responses. In such cases, one has to iterate again to identify the suitable questions and customers.

The next key step is to learn a Bayesian network model that *approximates the segmentation model*. All the questions in the survey questionnaire, and the segment are represented as nodes in this Bayesian network. Learning a Bayesian network model involves two steps: (1) structure learning—a Bayesian network is represented by a directed acyclic graph (DAG). The DAG structure could be learned with either score-based approach or constraint-based approach. The score-based approach first defines a criterion to evaluate how well the Bayesian network fits the data, for example, BIC score, then searches over the space of DAGs for a structure with maximum score [12, 13]. The constraint-based case uses the independence test to identify a set of edge constraints for the graph and then finds the best DAG that satisfies the constraints [14, 15]. (2) Parameter learning—this involves learning the parameters that are required to estimate the conditional probability tables of each node in the Bayesian network. These parameters are typically learned through expectation maximization, maximum likelihood, and gradient-based approaches. We use 70% of the survey data to learn the Bayesian network model.

A key advantage of a Bayesian network model is its ability to handle partial information at the time of inference, that is, the same Bayesian network model could be used for segment assignment even when fewer questions are asked. The main novelty in our approach is to exploit this property of Bayesian networks to reduce the number of questions in the survey. We find an optimal hyper parameter k , which is the number of random questions that could be asked to the customer whose responses, when fed to the Bayesian network model, will guarantee an average f-score above a configured threshold, for example, 0.70. In simple terms, how many fewer questions I could ask without compromising too much on the Bayesian network segmentation model performance. The algorithm that we used to identify the optimal number of minimal questions is shown in Fig. 56.3.

```

function FIND_K()
   $K \leftarrow [5, 10, 15, \dots]$             $\triangleright$  line search to find optimal value for  $k$ 
   $bestFscore \leftarrow 0.0$ 
   $bestK \leftarrow None$ 
  for  $k \in K$  do
     $segments \leftarrow dict()$ 
    for each customer in  $testSet$  do
       $questions \leftarrow randomQuestions(k)$ 
       $responses \leftarrow surveyResponses($ 
[9]  $questions, customers)$ 
       $segments[customer] \leftarrow bayesianNetworkModel($ 
[11]  $responses)$ 
    end for
     $metric \leftarrow fscore(segments)$ 
    if  $metric > bestFscore$  then
       $bestFscore \leftarrow metric$ 
       $bestK \leftarrow k$ 
    end if
  end for
  return  $bestK$ 
end function

```

Fig. 56.3 Algorithm to find an optimal number of minimum questions

56.2.2 Scaling Phase

Once the optimal value for k has been identified as explained in the previous section, the scaling phase becomes very simple. A customer is asked only k random questions, instead of going through the whole questionnaire. The responses to these k questions are passed through the Bayesian network model and the segment assignment is done. This approach also provides an opportunity for incrementally updating the segment assignment as new information becomes available. For example, the customer can also be questioned in multiple parts and the segment assigned to the customer can be updated based on his additional responses.

56.3 Results

We implemented our proposed approach to scale the market research that was performed for an Internet Service Provider (ISP) business. A total of 100,000 customers participated in the survey. The survey participants were sampled from the total customers based on their plan and lifetime value in a stratified manner. Most of the survey questions are scale based (1–5); a response of 1 means the participant strongly disagrees with the statement in the question, whereas a response of 5 means that the participant strongly agrees with the statement. A complete list of survey questions is shown in Table 56.2. The survey responses were analyzed, and a

Table 56.2 Survey questionnaire

ID	ABBR	Expansion	Actual question
1	AGP	Age group	
2	MAR	Marital status	
3	PAM	Perception about mobile	Find new technology exciting and want to have a mobile phone with the latest services and features.
4	AIE	Access to internet everywhere	It's important for me to be able to access the internet wherever I am
5	MAP	Most advanced products	I'm constantly looking for the most technologically advanced products available
6	DUT	Difficulty in using technology	For me to use a new technology product, somebody has to show me how to use it
7	TA	Technology avert	Feel that I am able to manage without many of the technology products that other people find essential
8	FVP	Features vs price	The features are more important than the price
9	U2D	Up to date	It is important to be up to date on major news
10	TFS	Technology for show-off	Carrying the latest technology products makes a good impression
11	U2P	Unwilling to pay	Even when I can afford them, I'm not willing to pay much for new technology products or services
12	DNB	Don't need mobile	Do not need a mobile phone
13	MBROW	Mobile browsing	Mobile browsing of the internet
14	MEMAIL	Mobile email	Send and receive email via the mobile phone
15	MBANK	Mobile banking	Mobile banking via the mobile phone
16	MVID	Mobile video	Watching videos on your mobile phone
17	GPS	Global position tracking	Mapping, navigation, or positioning service (like gps) via the mobile phone
18	GAM	Gaming	Playing video games is one of my favorite activities
19	SMP	Small payments	Small payment service via the mobile phone
20	TFF	Time for family	Spend a lot of time with my family
21	RURB	Rural or urban	
22	ELS	Life stage	
23	DIS	Diverse internet services	Derived attribute
24	SGV2	Segment labels	

combination of rule- and algorithm-based segmentation model was built, and four customer segments (S1, S2, S3, and S4) were identified.

In preparatory phase described in our approach, we learnt the structure of the Bayesian network using Hill Climbing (hc) greedy search on the space of directed graph and Akaike Information Criterion (AIC) as the scoring criteria. We used the Maximum-Likelihood estimates for fitting the parameters of the Bayesian network. For both structure learning and parameter fitting, we used the implementation available in the bnlearn R package. Figure 56.4 shows the structure of our Bayesian network model.

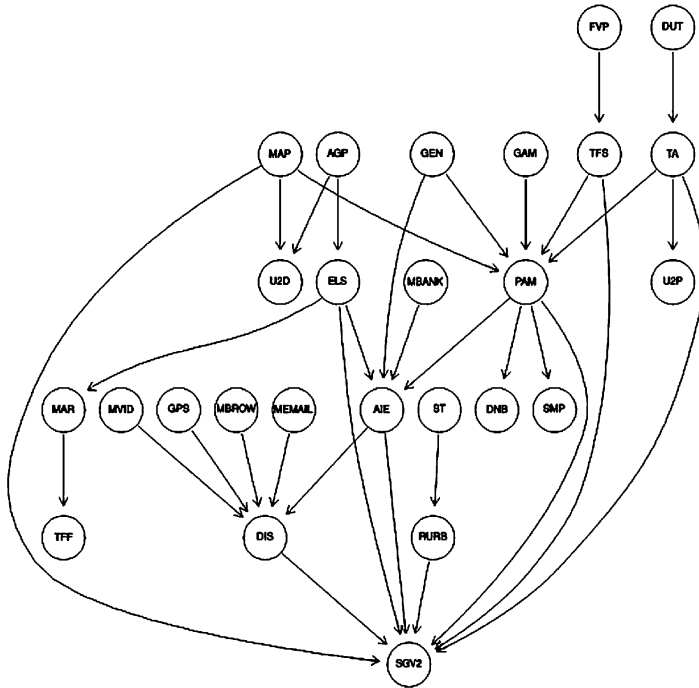


Fig. 56.4 Ask less – Bayesian network model

We used 70% (70,000) of the survey responses to learn the Bayesian network model. Note that the nodes in the model are responses to survey questions and the corresponding segment assignment for the customer (SGV2). The learned network structure was validated with domain experts, and we list few interesting observations: (1) A person’s perception about mobile (PAM) influences if he wants to access internet everywhere (AIE). (2) The final segment assigned to the customer is based on the fact if that customer uses diverse Internet services (DIS). (3) The gender of the customer (GEN) could influence the customer’s perception about mobile (PAM) and his urge to access Internet everywhere (AIE). (4) The customer’s value for features in a product (FVP) decides if he wants to use the product to showoff (TFS).

We used the line search algorithm shown in Fig. 56.3 to identify the optimal hyper parameter k , which is the number of random questions that could be asked to the customer that will guarantee an average f-score above 0.70. We used 30% of the survey responses (30,000) for this purpose. We have a total of 22 questions in the survey. We ran the *Find_k* algorithm with values for k as [5, 10]. The segment classification performance metrics of Bayesian network model for each value of k is shown in Tables 56.3, 56.4, and 56.5. We use the *cpquery* function of *bnlearn* to supply the partial evidence, that is, responses for randomly selected questions to run a conditional probability query and predict the segment assignment. We found the optimal value for k is 10 in this case. This means that by using our approach, we

Table 56.3 Accuracy metrics for $k = 20$

Segment	Precision	Recall	F-Score
S1	0.87	0.90	0.89
S2	0.68	0.79	0.73
S3	0.85	0.86	0.86
S4	0.96	0.81	0.88
Average	0.86	0.85	0.85

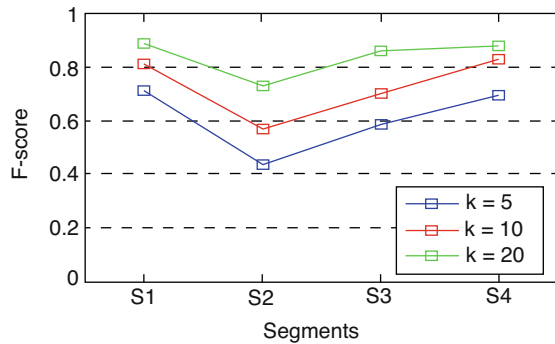
Table 56.4 Accuracy metrics for $k = 10$

Segment	Precision	Recall	F-Score
S1	0.77	0.87	0.81
S2	0.52	0.64	0.57
S3	0.76	0.65	0.70
S4	0.91	0.76	0.83
Average	0.75	0.74	0.74

Table 56.5 Accuracy metrics for $k = 5$

Segment	Precision	Recall	F-Score
S1	0.67	0.76	0.71
S2	0.38	0.52	0.44
S3	0.65	0.53	0.59
S4	0.79	0.63	0.70
Average	0.65	0.62	0.63

Fig. 56.5 Comparison of f-scores for different “ k ” values across segments



could reduce the number of questions by 50%. Figure 56.5 shows the comparison of scores for various values of k .

In the scaling phase, we integrate our Bayesian network model with the survey tool, which randomly selects k (10) questions and collects the responses for them from the customers. These responses are passed as evidence to the Bayesian network model, and segments are assigned.

56.4 Conclusion

In this chapter, we propose a simpler way to reduce the number of questions in a Market Research survey using Bayesian networks. We evaluated the effectiveness of our approach in a real-world setting, and we observe that our approach can help to reduce up to 50% of the questions with a minor dip in classification performance. Our work shows that Bayesian networks can serve as a simpler alternative to factor analysis to reduce the number of questions in a survey, without compromising the ability to collect information about various topics.

Acknowledgments We thank Prasad Garigipati, Henrik Palson, Andreas Timglas, and Roy Ollila for their help and support. Both the authors were introduced to the area of Market Research during their tenure at Xoanon Analytics. The value in asking fewer questions in a Market Research Survey was recognized by the authors based on their practical experience.

References

1. Wind Y, Douglas SP (1972) International market segmentation. *Eur J Mark* 6(1):17–25
2. Bradburn NM, Sudman S, Wansink B (2004) Asking questions: the definitive guide to questionnaire design – for market research, political polls, and social and health questionnaires. Wiley, New York
3. Cremonesi L (2016) High definition customers – a powerful segmentation. White paper, Ipsos MORI
4. Andrew Z, Peter D (2011) A guide to getting the best out of your segmentation analyses
5. Fricker Jr, Kulzy W, Appleget J (2012) From data to information: Using factor analysis with survey data, pp 30–34
6. Ehrenberg ASC, Goodhardt GJ S. I. M: Factor analysis: limitations and alternatives. Marketing Science Institute, Cambridge
7. Beri G (2007) Marketing research. Tata McGraw-Hill Education, New Delhi
8. Karvanen J, Rantanen A, Luoma L (2014) Survey data and Bayesian analysis: a cost-efficient way to estimate customer equity. *QME Quant Mark Econ* 12:305–329
9. Constantinou A, Fenton N, Marsh W, Radlinski L (2016) From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support. *Artif Intell Med* 67:75–93
10. Toyinbo P, Vanderploeg R, Belanger H, Spehar A, Lapcevic W, Scott S (2017) A systems science approach to understanding polytrauma and blast-related injury: Bayesian network model of data from a survey of the Florida National Guard. *Am J Epidemiol* 185(2):135–146
11. Salini S, Kenett R (2009) Bayesian networks of customer satisfaction survey data. *J Appl Stat* 36(11):1177–1189
12. Friedman N, Murphy K, Russell S (1998) Learning the structure of dynamic probabilistic networks. In: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, pp 139–147
13. Heckerman D, Geiger D, Chickering D (1995) Learning Bayesian networks: the combination of knowledge and statistical data. *Mach Learn* 20(3):197–243
14. Steck H (2001) Constraint-based structural learning in Bayesian networks using finite data sets
15. de Campos C, Ji Q (2011) Efficient structure learning of Bayesian networks using constraints. *J Mach Learn Res* 12(Mar):663–689

Chapter 57

Preferential Resource Selection and Scheduling of Cloud Resources Pivot on Value of Information



Renu Suresh Ganvir, Salaja Silas, and Elijah Blessing Rajsingh

Abstract The selection of resources and scheduling in the cloud are crucial due to the involvement of various features. Scheduling an appropriate resource onto the cloud is influenced by quality of service parameters. Providing a relevant resource to the user consists mainly of three steps: (1) finding the feasible set of resources, (2) selecting the most appropriate resource from the practical set of resources, and (3) scheduling the resource to the relevant processor. Selecting a relevant resource is modeled as a multi-criteria decision-making problem. Factors like availability, trust, cost, responsiveness, reliability, and capability have effects on the resource selection. In this chapter, an efficient workflow has been put into suggestion in consideration to make a selection of the most significant resource using PROMETHEE methodology, and scheduling is performed using a non-pre-emptive priority algorithm. The choice of the optimal resource is done pivoted on the value of information that is requested by the users for all the influencing factors. The outcome of the simulation proves that the suggested workflow decreases the response time, makespan, and cost, which also maximizes the quantity of resources utilized before the deadline.

Keywords Multi-criteria decision-making · resource selection · resource scheduling · PROMETHEE · Non-pre-emptive priority scheduling

R. S. Ganvir
Wipro Limited, Coimbatore, Tamil Nadu, India

S. Silas (✉)
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

E. B. Rajsingh
Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

Abbreviations

BaTS	Budget constraint scheduling
ELECTRE	Elimination and choice expressing reality
FCFS	First come, first served
IaaS	Infrastructure-as-a-service
MLBMPSO	multi-objective load balancing mutation particle swarm optimization
NBS	Nash bargaining solution
PaaS	Platform-as-a-service
PROMETHEE	Preference ranking organization method for enrichment of evaluations
RBS	Raiffa bargaining solution
SaaS	Software-as-a-service
SLA	Service-level agreement
VoI	Value of information

57.1 Introduction

Recently, cloud computing has an enormous impact on day-to-day life in providing services required by a user. Cloud is analyzed on service basis, for example, as IaaS, PaaS, and SaaS. Infrastructure-as-a-Service (IaaS) can configure and manage users' software such as operating system, middleware, and application. In platform-as-a-service (PaaS) users are allowed to use cloud setup for development, testing, delivery, and management of the cloud applications. Software-as-a-service (SaaS) allows the cloud services such as software delivery on demand. Cloud has different deployment models: public cloud, private cloud, and hybrid cloud. Public cloud is fully managed by the service offered by a third-party vendor. Private cloud consists of resources pertaining to cloud computing, which are used by a single organization, whereas hybrid cloud combines the features of both private and public clouds by the process or methodology that assists the sharing of data applications among them.

Many users frequently use resources that are offered by the cloud. As a result, when many providers tend to provide similar resources in the cloud with the same or different specifications, users may face difficulty in deciding which one to prefer pivot on their requirement. The preference also varies from one user to another. Therefore, selection of a resource should be pivot on the users' requirement and their preferences. An extensive quantity of resources is deployed to the same processor, and hence, scheduling resources on pivot of users' requirement onto the cloud is essential. Therefore, by allocating resources and scheduling, it is a major challenge influenced by resource availability, resource cost, reliability, capability, trust, and responsiveness is termed as value of information (VoI). Resource is selected to have good efficiency, PROMETHEE methodology has been executed for the cloud

service. Simulation is done on the cloud environment, and the resources are given to the selected processor. Scheduling resources in every processor is performed on priority pivot on consideration to improve efficiency in terms of cost utilization done on the resources.

In Sect. 57.2, we discuss the related work. The standard factors affecting resource selection in cloud are explained in Sect. 57.3. The suggested workflow for selecting resources and scheduling them in the cloud is explained in Sect. 57.4. The determination of the optimal resource using the PROMETHEE method is explained in Sect. 57.5. Resource scheduling using the non-pre-emptive priority method is discussed in Sect. 57.6. The simulation results obtained are analyzed in Sect. 57.7. The remarks of conclusion are discussed in Sect. 57.8.

57.2 Related Work

An elaborated literature survey has been done to understand the issues involved in resource selection and scheduling in cloud. To provide efficient resource selection and scheduling in cloud, many researchers have focused on various mechanisms for the selection of appropriate service providers and resource scheduling algorithms, in recent years.

Yong Dong (2010) proposed a resource scheduling approach considering two factors: user priority and length of the task. Using these factors, the tasks were assigned credits for scheduling [1].

Antony T et al. (2015) proposed min-to-min algorithm to minimize the makespan of the task by contemplating the length of task and the user's priority. The min-to-min algorithm considered the task length as chunks required for execution of tasks. The higher priority task is not given any special importance [2].

Awad AI et al. (2015) proposed multi-objective load balancing mutation particle swarm optimization (MLBMP SO) to achieve reliability by considering resource availability and perform rescheduling the task that failed allocation. The algorithm considered only few factors like reliability, execution time, and transmission time [3].

Weiwei Kong et al. (2016) considered auction method for scheduling VM in cloud. Factors like auction deadline and network throughput were taken into consideration. The cost was calculated pivot on the cloud attributes and VM configuration. Based on the output of auction, the tasks users requested were assigned to the opted VM resources [4].

Iyer GN and Veeravalli B (2011) suggested a pricing-based resource scheduling algorithm. Two parameters, namely, (1) Nash bargaining solution (NBS) and (2) Raiffa bargaining solution (RBS), were developed to obtain a favorable solution for the allocation of virtual CPU. NBS ensures proportional fairness, and RBS is responsible for task arrivals alongside task dynamics in real time. Asymmetric bargaining was attempted to reduce the cost without having an effect on

performance. This concept allowed the cloud user to opt different constraints such as resource deadline or resource budget for feasible resource allocation [5].

Opreescu A and Kielmann T (2010) proposed budget constraint scheduling algorithm (BaTS). BaTS schedule tasks into the cloud on the performance of the CPU and cost minimization of the completion time. The BaTS algorithm prioritized the user-defined budget and scheduled the task within a fixed budget [6].

Jain N et al. (2012) proposed market-based resource allocation. The model prioritized the job deadline rather than considering the servers that are queued into it, for which the CC-kcss-1 approximation algorithm was used. The deadline and the pricing were considered to transform allocation into the truthful mechanism. The algorithm was derived to maximize the interest using Bayesian assumption for cloud computing [7].

Pawar CS et al. (2012) proposed an algorithm for pre-emptible task execution and the service-level agreement (SLA) features for the likes of memory usage, network bandwidth, and the CPU utilization time to enhance resource utilization. The algorithm dynamically responded to the fluctuating workload by pre-empting the running execution task [8].

Silas S et al. (2012) proposed ELECTRE methodology as a multi-criteria decision analysis method for selecting the resources in cloud. The algorithm has considered the users' requirements that influenced the selection of the services in cloud [9].

Ladislau Bölöni and Damla Turgut (2017) proposed a computation mechanism, considering the monetary expenditure of the computation and the predicted financial benefits of the outturn that was considered as the value of information (VoI) [10].

In the literature, it has been identified that the value of information plays a foremost role in the administration of resources and resource scheduling with efficiency.

57.3 Standard Factors Affecting Resource Selection

From the literature, certain factors can be recognized, such as cost, responsiveness, reliability, availability, etc., which influence the determination of the resources onto the cloud. Many resource providers in cloud provides homogeneous resources but with different VoIs. User preferences and requirements also vary. In this section, the factor that affects the process of selecting the resources in cloud is put into discussion.

Resource Availability Current accessibility of resources, pivot on the requirements of the users [11].

Resource Cost It is computed by the time period required to implement the request, times the number of instructions executed [12].

Reliability It guarantees the performance according to its specifications [13].

Capability It is the amount of request that a resource can handle at a specific instance of time [14].

Trust It is taken as a five-level value from 1 to 5, where 1 is the least trust and 5 is the maximum trust. It is updated as per the feedback provided and heuristic information [14–16].

Responsiveness The capability of the resource provider enabling it to react to the users’ requirement quickly without introducing instability [17].

57.4 Proposed Workflow

The workflow for resource selection and scheduling is shown in Fig. 57.1. It contains standard modules like user registration, resource selection, monitoring system, and resource scheduling. The information is stored in a resource database and a resource provider database. Registering and allocation of resources is pivot on the users’ requirement and preferences.

Resource provider database contains and provides information about the available resources, and resource database contains the information of resources requested by users’ and the standard preference of each factor. The user request module permits users to request and utilize the resources provided by the cloud service. In accordance with users’ requirements, the resource selection module helps in recognizing the optimal resource from the resource provider. Resource scheduling module schedules the resources, pivot on the optimal scheduling algorithm, which

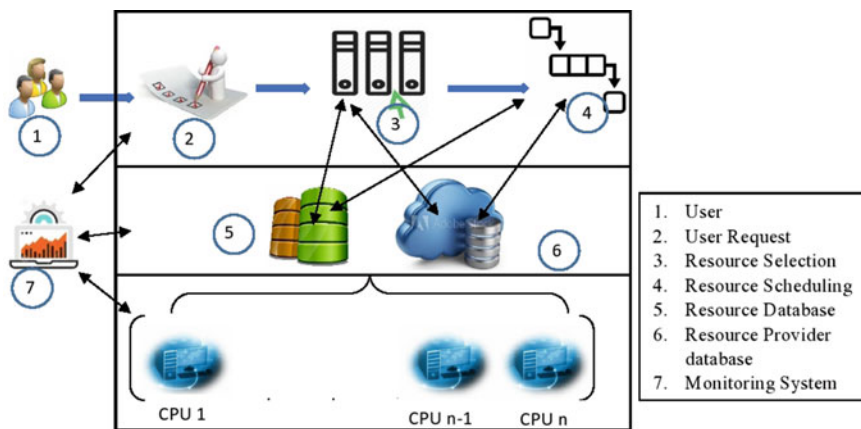


Fig. 57.1 Resource selection and scheduling workflow

helps the user to get the required resource in estimated time. Monitoring module monitors the resources allotted and its completion. Also, it updates the trust degree of the resource.

57.5 Resource Selection Algorithm

Resource selection is defined as the process of selecting the optimal resource from the set of finite resources that the resource provider provides, pivot on the requirements of the user and the preferences to each standard factor. The research motivation is to identify the best and feasible set among the finite set of the resources. Resource selection is influenced by many standard factors, and hence, it leads to NP-Hard Problem, which is further categorized as a decision-making problem dependent on multiple criteria.

PROMETHEE, as a preferential outranking methodology, is implemented to select the optimal resource from the service provider on the basis of users' requirement [18]. PROMETHEE gives privilege to the user to opt the resources as per their requirements and also give them privilege to enter the preferences for each factor.

Step 1: Users' Requirements

The set of outstanding preference $\{O_i, i = (1, 2, \dots m)\}$ for each user-provided factor F, is tabulated in Table 57.1.

for $i = 1$ to n :

$$\text{sum} += O[i],$$

for $i = 1$ to n :

$$O [i] = \frac{O [i]}{\text{sum}}$$

where summation of outstanding preferences is standardized to 1

$$\sum_{i=1}^m O_i = 1 \tag{57.1}$$

Table 57.1 Preference table

F_1	F_2	\dots	F_i	F_{i+1}	\dots	F_m
O_1	O_2	\dots	O_i	O_{i+1}	\dots	O_m

Step 2: Instance from Resource Provider

Information of related resources from resource and resource provider database is retrieved in order to generate ranking table, as per user’s requirements.

Table 57.2 shows data based on the user requirements with all the optimal alternative resource, A. n feasible resources are selected from the statistics available in the database.

Step 3: Estimation of Preference Function

The preferential function of each alternative resource is calculated for either maximization factor or the minimization factor. For maximization factor, the preferential function $Pf_j(A_x, A_y)$ gives the preference of alternative A_x over the alternative A_y ,

for $x = 1$ to p :

for $y = 1$ to p :

for $o = 1$ to m :

$$M[x] = (A[x]-A[y])*O[o].$$

Represented as,

$$Pf_j(A_x, A_y) = F_j[M_j(A_x, A_y)] \tag{57.2}$$

When the difference of preference function is negative, the function equals 0, such that

$$Pf_j(A_x, A_y) > 0$$

The calculated preferences for minimization factor,

$$Pf_j(A_x, A_y) = F_j[-(M_j(A_x, A_y))] \tag{57.3}$$

where $Mx_j(A_x, A_y) = F_j(A_x) - F_j(A_y), \forall A_j \in A$

Table 57.2 Ranking table

A	F ₁ (.)	...	F _j (.)	...	F _m (.)
A ₁	F ₁ (A ₁)	...	F _j (A ₁)	...	F _m (A ₁)
.
.
.
A _i	F ₁ (A _i)	...	F _j (A _i)	...	F _m (A _i)
.
.
.
A _n	F ₁ (A _n)	...	F _j (A _n)	..	F _m (A _n)

Step 4: Estimation of Overall Preference

In order to define index to overall preferences, pair-wise comparison is done from the selected workable set of resources.

Consider $A_x, A_y \in A$, where A is a set of substitute resources provided by the resource provider. Overall preference indices are calculated by

$$\theta(A_x, A_y) = \sum_{j=1}^m Pf_j(A_x, A_y)O_j \tag{57.4}$$

$$\theta(A_y, A_x) = \sum_{j=1}^m Pf_j(A_y, A_x)O_j \tag{57.5}$$

$\theta(A_x, A_y)$ gives the degree in which alternative A_x is preferred over the alternative A_y ; overall, the resource factors and $\theta(A_y, A_x)$ give the degree where the alternative A_y is preferred over alternative A_x for a specific service provider.

Step 5: Calculation of Outranking

Each alternative resource A_i is compared with remaining $n-1$ alternatives. The outranking is derived as

for $i = 1$ to n :

for $j = 1$ to n :

$$\text{sum1} = \text{sum1} + c[i][j].$$

$$\text{sum2} = \text{sum2} + c[j][i].$$

$$\text{positive}[i] = \text{sum1}/(n-1).$$

$$\text{negative}[i] = \text{sum2}/(n-1).$$

where sum1 is the positive outranking calculated as,

$$\lambda^+(A_i) = \frac{\sum_{x \in A} \theta(A_i, A_x)}{n - 1} \tag{57.6}$$

and sum2 is the negative outranking calculated as,

$$\lambda^-(A_i) = \frac{\sum_{x \in A} \theta(A_x, A_i)}{n - 1} \tag{57.7}$$

Step 6: Overall Outranking λ and Optimal Resource Selection Calculation

The overall outranking is derived as:

for $i = 1$ to n :

$$\text{Outranking}[i] = \text{positive}[i] - \text{negative}[i].$$

Mathematically represented as

$$\lambda(A_i) = \lambda^+(A_i) - \lambda^-(A_i) \tag{57.8}$$

The overall outranking is used to sort similar resource alternatives allocated by different service providers.

The estimated net outranking has the maximum $\lambda (A_i)$ is selected as the optimized resource alternative out of the viable set of alternatives, pivot on the requirements and preferences of the users.

Quantity of resources given to the specific processor is more, and hence, scheduling of resources is required to deliver the resource in optimal time.

57.6 Resource Selection Algorithm

Scheduling resources is an issue in the cloud environment, where there are certain resources that have to be executed in minimal time with maximum processor utilization. The factors that are under the influence of the processing time of a resource are the characteristics of each resource such as resource length, type of instructions, etc. [19] Scheduler schedules different resources, which are designated to the processor pivot on the scheduling algorithm. The scheduling algorithms can be put into categories as follows: pre-emptive and non-pre-emptive. In the pre-emptive scheduling algorithm, the process can go through interruptions by other prioritized resource and then resumed. In the non-pre-emptive algorithm, the resources given to the processor are put into execution with no interruption.

Non-pre-emptive priority-based scheduling is adapted for the complete resource's utilization without any interruption of the other process.

Step 1: Assignment of Priority to the Resources

Priority to resources is given pivot on the deadline of the resource.

$$\forall Pr_i, i = 1 \dots n$$

Here, n is the quantity of processors (number) Pr , R_i an ordered set of k resources pivot on the deadline represented by,

$$R_i = \{R_{i1}, R_{i2} \dots R_{ik}\} \text{ for the } i - \text{th processor}$$

If Deadline (R_{ip}) = Deadline (R_{iq})

then R_{ip} and R_{iq} are ordered pivot on the arrival time.

The resource in R_i is assigned with ordered priority

$P_i = \{P_{i1}, P_{i2} \dots P_{ik}\}$, such that P_{i1} is taken as the greatest priority, whereas P_{ik} has the least priority.

Step 2: Resource Scheduling on the Basis of Priority Algorithm

The ordered resources are scheduled, pivot on the non-pre-emptive priority scheduling algorithm, the completion time (CT), the wait time (WT), and the burst time (BT) are calculated as follows:

$$CT_{ij} = WT_{ij} + BT_{ij}, j = 1 \dots k \tag{57.9}$$

where k is the quantity of resources (number) scheduled to the particular processor Pr_i ,

$$WT_{ij} = CT_{ij-1} \tag{57.10}$$

57.7 Experimental Studies

The proposed methodology was simulated in Cloudsim, and the experimental results were obtained. The simulation parameters are tabulated in Table 57.3.

Experiment 1 The users are permitted to register their requirements and the preferences in cloud. Each resource provider was permitted to offer 5, 10, 15, 20, and 25 resources. To detect the best resource from the finite set of resources that the service provider provides, on the basis of six criteria and the job requests, it is 50 at a particular time t .

Figure 57.2 shows the variation in the service selection using the PROMETHEE and ELECTRE methods. It is observed that the PROMETHEE-based resource selection takes minimum time to select resource compared to ELECTRE due to its inherent characteristics.

Table 57.3 Simulation parameters

Parameter	Values
Load	50%
Resources	0–50
Simulation time	0–500 s
No. of resources	0–100
No. of criteria	0–6

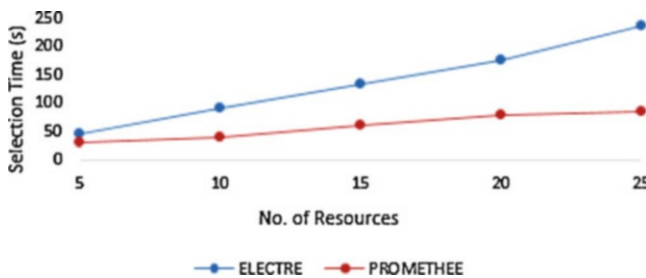


Fig. 57.2 Selection of the best resource

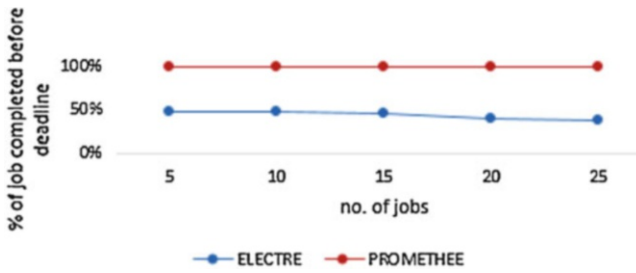


Fig. 57.3 Percentage of jobs completed before deadline

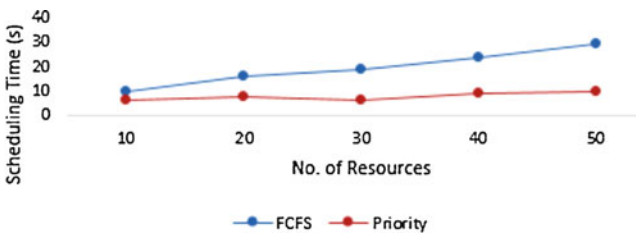


Fig. 57.4 Resource scheduling

Experiment 2 The users are permitted to register their requirements in the cloud such as the preference values for the factors. Each resource provider was permitted to offer 5, 10, 15, 20, and 25 resources. Detection of the best resource from the limited set of resources provided by the service provider is dependent on the six criteria, and 25 resources are considered.

It is observed in Fig. 57.3 that the percentage of job completion is high in PROMETHEE when compared with the ELECTRE method for the selected resources.

Experiment 3 To organize the resources in CPU through scheduling, the resources are varied as 10, 20, 30, 40, and 50, assigned to the processor and scheduled using non-pre-emptive priority and FCFS, respectively. In Fig. 57.4, it is perceived that priority scheduling takes minimum time to utilize resources in processor than in FCFS.

Experiment 4 To schedule the resources in CPU, the resources are varied as 10, 20, 30, 40, and 50, assigned to the processor and scheduled using non-pre-emptive priority and FCFS, respectively. In Fig. 57.5, it is perceived that the makespan of priority is less than the FCFS.



Fig. 57.5 Makespan of the job

57.8 Experimental Studies

In accordance to opt for the optimal resource and to schedule resource, pivot on the users' requirements the decision has to be made by the cloud provider. Best resource selection is dependent on multiple criteria as a decision-making problem. In this chapter, we have discussed regarding the resource selection problem and resources scheduling with different factors, which influence to select the resources and the priority factor to each resource.

References

1. Dong Y (2010) Power measurements and analyses of massive object storage system. Computer and Information Technology (CIT); IEEE 10th international conference, pp 1317–1322
2. Antony T, Krishnalal G, Jagathy Raj VP (2015) Credit based scheduling algorithm in cloud computing environment. Proc Comp Sci 46:913–920
3. Awad AI, El-Hefnawy NA, Abdel_kader HM (2015) Dynamic multi-objective task scheduling in cloud computing based on modified particle swarm optimization. Adv Comput Sci Int J 4 (5, No.17):110–117
4. Kong W, Yang L, Ma J (2016) Virtual machine resource scheduling algorithm for cloud computing based on auction mechanism. Optik – Int J Light Electron Optics 127 (12):5099–5104
5. Iyer GN, Veeravalli B (2011) On the resource allocation and pricing strategies in Compute Clouds using bargaining approaches. 17th IEEE International Conference on Networks (ICON), pp 147–152
6. Opreacu A, Kielmann T (2010) Bag-of-tasks scheduling under budget constraints. IEEE second international conference on cloud computing technology and science (CloudCom), pp 351–359
7. Jain N, Menache I, Naor J, Yaniv J (2012) Near-optimal scheduling mechanisms for deadline-sensitive jobs in large computing cluster. In: Proceedings of the twenty-fourth annual ACM symposium on parallelism in algorithms and architectures, pp 255–266
8. Pawar CS, Wagh RB (2012) Priority based dynamic resource allocation in Cloud computing. International Symposium on Cloud and Services Computing (ISCOS), pp 1–6
9. Silas S, Rajsingh EB, Kirubakaran E (2012) Efficient service selection middleware using ELECTRE methodology for cloud environments. Inf Technol J 11(7):868–875

10. Bölöni L, Turgut D (2017) Value of information based scheduling of cloud computing resources. *Futur Gener Comput Syst* 71:212–220
11. Galli G, Gebert AD, Otten LJ (2013) Available processing resources influence encoding-related brain activity before an event. *Cortex* 49:2239–2248
12. Carrasco RA, Iyengar G, Stein C (2018) Resource cost aware scheduling. *Eur J Oper Res* 269(2):621–632
13. Zhang Q, Cheng L, Boutaba Cloud computing: state-of-the-art and research challenges. *J Internet Serv Appl* 1(1):7–18
14. Ahamed SI, Sharmin M (2008) A trust-based secure service discovery (TSSD) model for pervasive computing. *Comput Commun* 31(18):4281–4293
15. Zhou J, Abdullah NA, Shi Z (2011) A hybrid P2P approach to service discovery in the cloud. *Int J Info Technol Comput Sci* 3:1–9
16. Wang Y, Vassileva J (2007) Toward trust and reputation based web service selection: a survey. *Int Trans Syst Sci Appl (ITSSA) J* 3(2):118–132
17. Wendell P, Jiang JW, Freedman MJ, Rexford J DONAR: decentralized server selection for cloud services. In: *Proceedings of the ACM SIGCOMM 2010 conference, New Delhi, India*, pp 231–242
18. Silas S, Rajsingh EB, Ezra K (2013) An efficient service selection framework for pervasive environments. *Int J Wirel Mob Comput* 6(1):80–90
19. Vickson RG (1980) Choosing the job sequence and processing times to minimize total processing plus flow cost on a single machine. *Oper Res* 28(5):115–167

Chapter 58

A Survey on Supervised and Unsupervised Learning Techniques



K. Sindhu Meena and S. Suriya

Abstract Supervised learning is the popular version of machine learning. It trains the system in the training phase by labeling each of its input with its desired output value. Unsupervised learning is another popular version of machine learning which generates inferences without the concept of labels. The most common supervised learning methods are linear regression, support vector machine, random forest, naïve Bayes, etc. The most common unsupervised learning methods are cluster analysis, K-means, Apriori algorithm, etc. This survey paper gives an overview of supervised algorithms, namely, support vector machine, decision tree, naïve Bayes, KNN, and linear regression, and an overview of unsupervised algorithms, namely, K-means, agglomerative divisive, and neural networks.

Keywords Supervised learning · Unsupervised learning · Support vector machine · Decision tree · Naïve Bayes · KNN · Linear regression · K-means · Agglomerative divisive · Neural networks

Abbreviation

KNN	K-nearest neighbor
WSD	Word-sense disambiguation
CNN	Convolution neural network
DT	Decision tree
NB	Naïve Bayes

K. Sindhu Meena · S. Suriya (✉)
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

58.1 Introduction (Fig. 58.1)

Supervised Learning Techniques: It involves two phases, namely, training phase and testing phase. This technique trains any system through a training phase, which consists of input data along with target results. After training phase, the system undergoes a testing phase using inputs without target labels. It is a crucial, fast, and accurate technique. Supervised learning is grouped into two divisions: regression and classification. Classification focuses on categorical variables like “black,” “white,” “yes,” and “no.” Regression focuses on variable that possesses real values like “money,” “height,” and “age” (Fig. 58.2)

Unsupervised Learning Techniques: It attempts to classify the input dataset into classes or clusters. The input dataset is a raw data without any class labels or target results. The raw data refers to unstructured dataset without class labels, optimization criteria, and feedback. The goal of unsupervised learning is to find natural partitions in the dataset set. Unsupervised learning techniques can be used for smart preprocessing and feature extraction systems (Fig. 58.3).

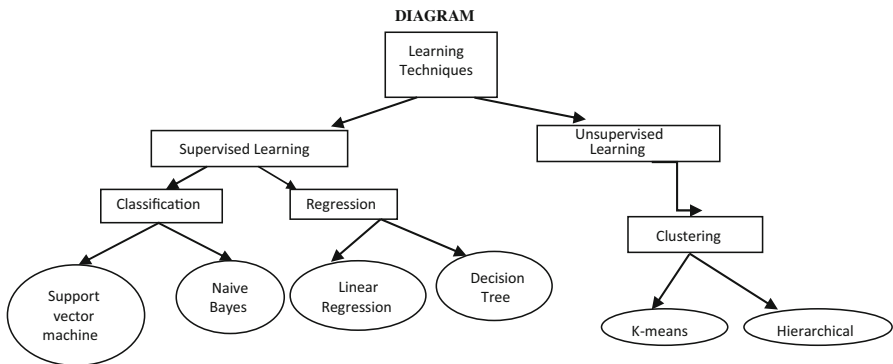


Fig. 58.1 The overall diagrammatic representation of learning techniques

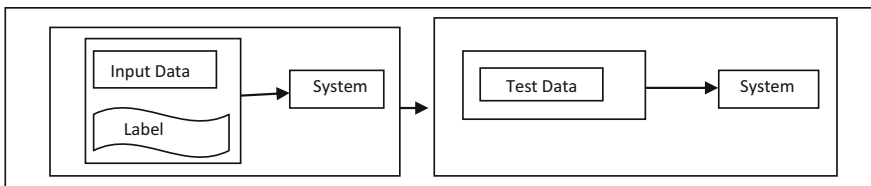


Fig. 58.2 The diagrammatic representation of supervised learning techniques

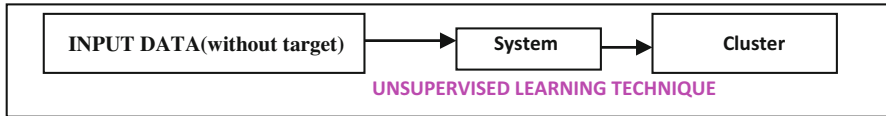


Fig. 58.3 Unsupervised learning techniques

58.2 Literature Survey

58.2.1 Supervised Learning

58.2.1.1 SVM

Paper [1] aims to promote research in sentiment analysis of tweets by providing annotated tweets for training, development, and testing. The objective of the system is to label the sentiment of each tweet as “positive,” “negative,” and “neutral.” They describe a Twitter sentiment analysis system developed by combining a rule-based classifier with supervised learning. Benefits of using this system are to maximize the positive and negative precision and recall; the rule-based classifier is used to correct or verify the neutral SVM predictions. This paper uses a modified version of SVM algorithm. Another paper [2] proposes the conception and implementation of computer-assisted detection (CAD) for the classification of mammographic images. It is to detect the breast cancer early using CAD. The application used here is Breast Cancer Classification. Benefits of using this system are to identify the breast cancer earlier. It classifies the nature of tumors in terms of benignity or malignancy. This paper uses a modified version of SVM algorithm. Paper [3] suggests an extraction of features for classifying the EEG signals into three stages. The first stage includes calculating the empirical mode decomposition (EMD) from the EEG signals and produces a set of intrinsic mode of function. To process further, we can make use of first three IMFs. In the next stage, the features are extracted by calculating the temporal and spectral attributes of IMF. The power spectral density is used for calculating the spectral attributes. The Hilbert-transformed IMF is used for fetching the temporal and spectral components. This transformation helps in removing the DC from the spectral attributes. The spectral signal is one of the main sources of nonstationary signals. The last stage makes use of support vector machine (SVM) for EEG signal classification. The application they have used here is for classifying the seizures and epilepsy in the EEG signals. The benefit of applying SVM classifier is to obtain the best accuracy. The basic algorithm is used here. Paper [4] describes an approach for automatic construction of dictionaries for named-entity recognition (NER). First, we collect a high-recall, low-precision list of candidate phrases from

the large unlabeled data collection for every named entity type using simple rules. The second step is to build the accurate dictionary of named entities by eliminating the noisy information from the list in the previous step. By using virus (GENIA) and disease (Dogan), the dictionaries are evaluated. It can be applied directly to the dictionary-based taggers. The advantage of this approach is different from binding belief about the data. SVD can be solved and better performance is achieved. It makes use of seed examples that are compiled manually. The modified version of SVM is used. Paper [5] aims to assess the neural predictors of long-term treatment outcome in participants with social anxiety disorder (SAD) 1 year after the completion of Internet-delivered cognitive behavioral therapy (CBT). SVM allows the predictions at the individual level, where it is trained to separate long-term treatment responders from nonresponders based on blood oxygen level-dependent (BOLD) responses to self-referential criticism. It can be predicted by using fMRI and SVM classification. The advantage of using SVM is that it provides accurate prediction of treatment outcome at the individual level. The modified version of SVM is used here. Paper [6] presents an effective method for gene classification using SVM. It solves the complex classification problems. Mutual information (MI) between the genes and the class labels is used for identifying the informative genes. The selected genes are utilized for training the SVM classifier, and the testing ability is evaluated using the leave-one-out cross validation (LOOCV) method. It reduces the dimensions of the input features by identifying the most informative gene subset and improves classification accuracy. This paper uses a basic SVM model. Paper [7] concentrates on the accuracy of least square support vector machine (LSSVM) and multivariate adaptive regression splines (MARS) and M5 model tree in identifying the river water pollution. The LSSVM and MARS provide the equal accuracy, and better performance is acquired compared to the M5 model tree. The monthly water pollution level from the river is obtained by using ammonia (AMM), water temperature (WT), and total Kjeldahl nitrogen (TKN) attributes. The application they have used is long-term prediction river water pollution. The modified version of SVM is used.

58.2.1.2 Decision Tree

Paper [8] explains the medical care quality and the status, development, and variation related to the quality by data. Data mining helps to find out the problems of management, various possible causes of the problem, and the corresponding solution strategies. There are two main methods: first one is using statistical methods to analyze a certain medical quality indicator, and another one is multidimensional comprehensive evaluation methods like rank sum ratio method and synthetic index method. The C4.5 algorithm helps to select hospital's inpatient homepage as training set and test set to realize the predictive parsing of hospital's medical quality indicator vividly by graphical means and finally calculates the accuracy of prediction by test set. The decision tree is the algorithm used. Paper [9] intended to demonstrate the data mining technique in the disease prediction systems in medical domain. The

heart disease-based data is selected for analysis and prediction. It gives patients a warning about the probable presence of heart disease even before he visits a hospital or tends toward a costly medical checkup. Benefits of using a decision tree are that it is more accurate than other methods and it classifies and predicts the heart disease. The decision tree algorithm is used here. Paper [10] uses the sentiment analysis to study the people's opinions, attitudes, and emotions toward an entity. Opinion helps to collect information about the positive and negative aspects of a particular topic. Finally the positive and highly scored opinions obtained about a particular product are recommended to the user. There are several challenges in sentiment analysis. The first is an opinion word that is considered to be positive in one situation may be considered negative in another situation. The second challenge is that people do not always express opinions in the same way. The basic model of decision tree is used here. Paper [11] focuses on the issues of leakage of information which impacts on both financial organization and customers. And it proposes a new approach that uses combined supervised learning techniques to classify the information in order to avoid releasing information that can be harmful for either financial services providers or customers. They entitled this approach as Supervised Learning-Based Security Information Classification (SEB-SIC). The goal will be achieved by using the supervised learning techniques to predict whether the information sharing will be hazardous for any relevant parties. The prediction process will be executed by decision tree-based risk prediction (DTRP). The modified version of DT is used here. Paper [12] describes the network traffic in the Internet, either based on the user demand for network resources, QoS scheduling, or development trend of network applications for expansion transformation of the existing network; here various network applications need to be classified and identified accurately. The benefit of using Hadoop platform is that it will improve the C4.5 decision tree algorithm parallelization. It improves parallel algorithm and has faster speed and higher accuracy. The modified version of decision tree is used here. Paper [13] attempts to propose a supervised approach decision tree for word-sense disambiguation task in Assamese language. This paper aims to disambiguate the words which have multiple sense in a context automatically. Some challenges like sense inventory along with their senses were discovered, and sense-annotated data as a training sample was manually prepared. Precision, recall, and F-measure were used as metrics for WSD. Assamese is a less computational-aware language, and WSD task using a supervised approach DT with cross-validation evaluation was the first initiative toward Assamese language. This will provide a remarkable contribution to NLP fields. The basic model of DT is used here. Paper [14] aspires an effective spam filtering technique related to the decision tree in the data mining technique. The spam is analyzed by using the association rules. The rules are applied for developing the systematized spam filtering method. Benefits are as follows: the revision information learned from the analysis of misjudged emails incrementally gave feedback to their method and its ability to identify spam would be improved. The basic model of DT is used here. In paper [15], a new supervised one-side sampling technique to preprocess the imbalanced dataset is used to overcome three main difficulties. First, the customer churn dataset is substantially imbalanced in reality. Second, the samples in

feature space are relatively scattering. Third, the dimension of feature space is high and dimension reduction is necessary for algorithm efficiency. C5.0 decision tree is the classifier applied in this study to predict customer churn in two or three months based on current information. Benefit of this method is that it retains the potential lost customer. The modified version of decision tree is used here.

58.2.1.3 Naïve Bayes

Paper [16] proposes a scheme for adopting the Laplace smoothing technique with binarized naïve Bayes classifier for enhancing the accuracy and employing SparkR for speed up via distributed and parallel processing. Binarized NBC (BNBC) was developed to solve this problem by not counting the frequency of each word but checking its existence. Their proposed approach effectively reflects this notion in the calculation of the probabilities of the classifier. The benefit of using this application is that if a word not existing in the training data appears, the probability of 0 occurs during the test of new document. This results in decreased accuracy of NBC. Laplace smoothing has thus been adopted to resolve this problem. In sentiment analysis, the presence of a word is more important than its frequency. Paper [17] proposes a simple, efficient, and effective feature weighting approach, called deep feature weighting (DFW), which estimates the conditional probabilities of naïve Bayes by deeply computing feature weighted frequencies from training data. The authors focus their attention on feature weighting approaches and find that all of the existing feature weighting approaches only incorporate the learned feature weights into the classification of formula of naïve Bayes and do not incorporate the learned feature weights into its conditional probability estimates at all. Naïve Bayes with our deep feature weighting rarely degrades the quality of the model compared to standard naïve Bayes and, in many cases, improves it dramatically. Finally, the paper applies the deep feature weighting to some state-of-the-art naïve Bayes text classifiers and has achieved remarkable improvements. In paper [18], the goal is to diagnose the diabetes types and the level of risk of the diabetic patients. The data mining techniques are used like clustering and classification. This paper claims in creating the expert clinical system to diagnose diabetes mellitus. The diabetes type and risk of the every patient can be analyzed by using the data mining techniques. Paper [19] aims to effectively detect Android malware based only on requested permissions. The authors explored machine learning techniques used to learn and train application profiles and to detect and predict application status: either malicious or normal. They evaluated and discussed the system with renowned antiviruses. Their accomplishments are good in detecting the accuracy. It is 98% accurate, and it predicts 96% of true positive rates. This means that it is capable of discriminating almost all cases of malware in detection and prediction. Paper [20] focuses on the impact of feature selection and engineering in the classification of handwritten text by identifying and extracting those attributes of the training dataset that will contribute the most toward the classification task using classifiers like J48, naïve Bayes, and sequential minimal optimization (SMO). This results in improved accuracy of the classifiers as

compared to the work reported earlier. The performance evaluation of the classifiers used for OCR and pattern recognition is done. Initial classification performance of all the classifiers listed above was recorded on the raw dataset. Finally, the dataset was transformed after performing relevant feature selection and engineering on its attributes. The same classifiers were again trained on the transformed dataset and their accuracy was recorded. This paper uses the widely used MNIST dataset of handwritten digits for training the classifiers. In paper [21], the goal is to develop techniques to continuously and automatically detect a smartphone user's mobility activities, including walking, running, driving, and using a bus or train, in real-time or near real-time (<5 s). Their application is sensing mobility contexts using smartphone sensors. They demonstrated that, by combining data from GPS, accelerometers, and GIS with existing ML algorithms, one can build a highly performing classifier for detecting mobility contexts of smartphone users. Thus, the computational complexity of the classification algorithms suggested that many of these classifiers can feasibly be implemented in a smartphone.

58.2.1.4 KNN

Paper [22] explores opinion mining using supervised learning algorithms to find the polarity of the student feedback based on predefined features of teaching and learning. It is for predicting the polarity of the student comments based on extracted features like examination, teaching, etc. It uses K-nearest neighbor for improving the performance. KNN classifier employs an indexing mechanism for the training datasets. To classify a document, it calculates the similarity of the document with the training set index and uses the k-nearest neighbor by measuring the similarity by functions such as Euclidean distance. It classifies the positive and negative based on their polarity to analyze the features which need improvement. The basic KNN model is used here. Paper [23] presents a genetic algorithm optimized K-nearest neighbor algorithm (evolutionary KNN imputation) for missing data imputation. This paper addresses the effectiveness of using supervised learning algorithm for missing data imputation. This paper mainly focuses on local approach where the proposed evolutionary K-nearest neighbor imputation algorithm falls in. The evolutionary K-nearest neighbor imputation is the extension of K-NN. The importance of using machine learning algorithm in this paper is that most of the common imputation methods such as case deletion and mean imputation method are showing less effective results by not considering the correlation of data. The mean error from mean imputations is not effective, and the error rate is relatively higher than the supervised algorithms. Paper [24] analyzes the groundwater potential mapping to get better and more accurate groundwater potential mapping (GPMs). The performance of K-NN is excellent. The paper concluded that spring occurrence had a direct relationship with TWI, while a reverse relationship was observed between spring occurrence and two factors including slope length and distance from faults. The importance of groundwater influence factors changes in different places and feature selection needs to be investigated before the modeling process. It can be used in

different areas for groundwater potential mapping. The modified version of K-NN is used here. Paper [25] uses different seismicity indicators as an input for system to predict earthquakes, which is becoming increasingly popular. The different attribute for the input in supervised learning technique is created by using the new techniques. It relates to training and testing the dataset for calculating the b-value. This method is applied to the four Chilean zones. It can be extended to any location. When the dataset length selection is good, the prediction accuracy is also good. The attributes of some indicators result in better accurate prediction. The basic model of KNN is used here. Paper [26] proposes an approach of multiple-implementation testing to test supervised learning software, a major type of ML software. This paper derives a test input's proxy oracle from the majority-voted output running the test input of multiple implementations of the same algorithm. K-NN algorithm is used here. This approach is highly effective in detecting faults in real-world supervised learning software. The majority-voted oracle has low false positives and can detect more real faults than the benchmark-listed oracle. The modified version of KNN is used. Paper [27] focuses to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to real profile within the same network by analyzing the content of the comments generated by both profiles. It presents a successful real-life use case in which the methodology is applied to detect and stop a cyberbullying situation in a real elementary school. The KNN algorithm helps to check the effectiveness thereof have been $k = 1-5$ to determine if taking account of more neighbors significantly improves the result. To optimize the result, filter the tweets text with stop words. It helps in identifying the real user or users behind a trolling account. The basic model of K-NN is used here. Paper [28] proposes the word embeddings which help to learn the semantic for words from a given sentence. The Word Mover's Distance (WMD) helps in finding the dissimilarity between the two text documents. The minimum distance of the embedded words of a single document need to travel for reaching the embedded word of other document. The WMD metric leads to unprecedented low k-NN document classification error rates. It discovers how quickly the lower bound distance helps to speed up the nearest neighbor. It is done by the collecting and pruning the neighbor. The advantage of using the WMD is to reduce the error rates in the datasets. It obtains the intelligence about the text documents in English language. It reinforces interpretability. The penalty for the terms is added if any two words take place in different selection of the same structured documents. The basic model of K-NN is used here. Paper [29] focuses on classifying the leukemia patients' dataset. In case of abnormal leukemia, it leads to bleeding, anemia, and impaired ability to survive against the infection. Biomarkers give the doctors the necessary information regarding the class of therapy required for the patient. It also provides the information about patients having severe condition of growing the disease. This paper proposes a research work on selecting biomarkers from gene-based microarray dataset and then predicting the subtype of cancer. K-NN is of higher accuracy. The basic model of KNN is used here.

58.2.1.5 Regression

Paper [30] proposes how a universal sentence is trained by using the supervised data of natural language inference datasets that exceed unsupervised methods like skip thought vectors. It is similar to the computer vision which uses the Image Net for acquiring the features that can be transferred to other tasks. It shows the suitable natural language inference for learning other NLP tasks. In a large corpus, the sentence encoder model is trained and transferred to other tasks. They solve the following two questions: Which is the preferable network architecture, and how is the task trained? The best results like accuracy are obtained while training the natural language tasks. In paper [31], the goal is to use two supervised methods like regression and classification. While learning from the crowds, the heterogeneity and biases among different annotators are discovered. The stochastic variational inference algorithm is used to scale a larger dataset. The proposed model joins the words in a document from the mixture of topics. The topics for the document are analyzed. The labels for the different annotators are a noisy version of the latent ground. Regression concentrate is used to handle the multiple annotators with different biases and reliabilities when the target variables are continuous variables. The altered version of regression is used. Paper [32] works toward helping the readers in transforming the motivation and formulating the problems and methods of the strong machine learning techniques for future networks in order to move into the unexplored applications. The Future 5G mobile will access in self-manner. It accesses the spectral band with the help of spectral efficiency learning and inference to control the transmission power and adjusting transmission protocols with the help of quality of service. The study uses regression algorithm. The goal of regression analysis is to predict the value of one or more continuous-valued estimation targets, given the value of a D -dimensional vector x of input variables. The estimation target is a function of the independent variables. Regression models can be used for estimating or predicting radio parameters that are associated with specific users. The basic model of regression is used. Paper [33] proposes a cross-modality consistent regression (CCR) model, which is able to utilize both the state-of-the-art visual and textual sentiment analysis techniques. They first fine-tune a convolutional neural network (CNN) for image sentiment analysis and train a paragraph vector model for textual sentiment analysis. On top of them, they train the multi-modality regression model. The results show that the proposed model can achieve better performance than the state-of-the-art textual and visual sentiment analysis algorithms alone. They would like to learn people's overall sentiment over the same object from different modalities of the object provided by the user. In particular, they focus on inferring people's sentiment according to the available images and the short and informal text. They used a modified version of regression model. In paper [34], trouble with learning the mixture of regression models for the individuals is analyzed. The mixing component is unknown, and it has to be

determined in the unsupervised method using data-driven tools. The novel penalized method is established to pick the mixing components, to know the mixture proportions, and also to know the unknown attributes in the models. It deals with the continuous and discrete response by initial two moment conditions.

58.2.2 *Unsupervised Learning*

58.2.2.1 **K-means**

Paper [35] proposes a simple and effective approach to automatically learn a multilayer image feature for satellite image scene classification. The feature extraction is composed of two layers which are uniformly learned by the K-means algorithm. It can also extract the complex features. In the multilayer approach, the feature from the higher layer has better classification performance than the lower layers. It achieves better performance compared with any single layer. It can achieve the VHR Image scene classification performance compared with the state-of-the-art approaches. K-means have a better performance when compared to SC and S-RBM. The benefit of this method is that it automatically mines the structure information from simple to complex, which meets the hierarchical perceptual mechanism of the human visual cortex. The basic model of K-means is used here. Paper [36] explores the application of the spherical k-means algorithm for feature learning from audio signals. They evaluate the approach on largest public datasets of urban sound sources available for research, and compare it to baseline approach by configuring it to capture the temporal dynamics of urban sources. It mainly focuses on the classification of auditory scene type (e.g., street, park), as opposed to the identification of specific sound sources in scenes such as car horn and bird tweet. It does not outperform the baseline approach. The modified version of K-means is used here. Paper [37] deals with the emotion recognition. It is to identify the emotions and assign one of the seven emotions to the short videos which are obtained from the Hollywood movies. The video clips represent the emotions based on the realistic condition by using the attributes like pose and illumination and explore by combining the feature from multiple labels. Deep learning techniques are used in this paper for focusing on modalities. CNN helps to capture the visual information by detecting the faces. The visual features around the mouth are extracted by using the K-means bag of mouth model. Better accuracy is achieved. Benefits are that it can be used in large-scale mining of imagery from Google image search to train the deep neural network. The basic model of K-NN algorithm is used. Paper [38] focuses on the unsupervised deep learning techniques to know the details of the patient from the electronic health data record (EHR) for clinical predictive model. The problems of supervised methods are overcome by using the unsupervised methods. The unsupervised methods help to identify the patterns automatically and make an overall representation which is easier to extract the needed information automatically while building the classifiers. To preprocess the level of the patient, the unsupervised

methods are used. It clusters the EHR data for understanding better. It uses deep patient for representing the general features of the patient which are obtained from the deep learning technique. K-means is the algorithm used here. K-means help to cluster the unlabeled data into k-clusters so that each point belongs to the cluster which has the closest means. The centroids of the cluster help to produce the features in future. The main focus of this paper is to assess the methods on EHR data warehouse for consolidating the results. A large number of patient results are also evaluated. The basic model of K-means is used here. Paper [32] works toward helping the readers in transforming the motivation, formulating the problems and methods of the strong machine learning techniques for future networks for the unexplored applications. The algorithm is done iteratively, and the object is allocated to the particular cluster where its centroid is close to the object which is based on the Euclidean distance. Clustering is the main problem for 5G network mainly while concentrating on the heterogeneous scenario. It is to reduce the traffic in the wireless system by utilizing the high-capacity optical infrastructure. It uses initial gateway access point which helps to pluck randomly from the set of MAP, and it intelligently determines the meritorious initialization criterion. Each MAP is allocated to the nearest GAP. If many GAPs are in the environment, then the GAP which has the virtual channel will be chosen. By using the classic K-mean clustering, the MAPs are split into k groups with the close GAPs. The basic model of k-means is used.

58.2.2.2 Agglomerative

Paper [39] proposes the unsupervised learning method to discover the groups of molecular system based on the similarity in the structure by using the Wards minimal variance objective function. The minimal variance clustering is applied to the set of tripeptides by using the informatics theory to find how point mutations affect the protein dynamics. Here it focuses on the ability of the unsupervised MVCA algorithm to identify groups of molecular systems. Hierarchical agglomerative clustering is a set of pairwise distances between data points and iteratively merges the two closest clusters or singletons. Here the agglomerative and similarity distance function is used to quantify the similarity between multiple models for related dynamic systems. This analysis is intended to address a knowledge gap specific to supervised ML for chemoinformatic analyses designed to predict the properties of novel molecules. The agglomerative algorithm is used here. Paper [40] focuses on the membrane computing. It uses the fuzzy membrane, computing techniques, and clustering algorithm. Fuzzy clustering algorithm achieves good fuzzy partitioning for a dataset. Each element as a separate cluster and merge them into two larger cluster, or divisive which begin with the whole set and successively divide it into small cluster. The membrane computing aims to abstract computing models from the structure and functioning of living cells. The main aim of this paper is to solve fuzzy clustering problems. It uses an optimal cluster for a dataset and determines good fuzzy portioning. The basic model of HIAL is used here. Paper [41] focuses on the possibility of remote monitoring and screening of Parkinson's disease and

age-related voice impairment for the general public using self-recorded data on readily available or emerging techniques such as smart phones or IoT devices. It uses a sustained Vowel/a/recordings using iPhone. The purpose of finding the number of clusters is not just trying to find the number of optimal K to best reflect the level of impairments, but also to see how well the subjects can be clustered. The AGNES method joins one closest pair of clusters in each iteration by comparing the distance. The AGNES algorithm performs the best. This study has achieved clustering the voice with a data reduction ratio of 518. The basic model of AGNES is used. Paper [42] explains how attacker track mobile user is explained. Because users carry their phone everywhere where some sensitive or private information is stored about individual user. As user interacts with mobile apps, lot of network traffic is generated by sending or receiving request. Here they try to act as an eavesdropper which attacks on network traffic of the device from controlling Wi-Fi access point. Paper [43] explores the use of unsupervised clustering based on passive DNS records and other inherent network information to identify domains that may be part of campaigns but resistant to detection by domain name or time-of-registration analysis. They found that using this method can achieve up to $2.1 \times$ expansion from a seed of known campaign domains with $<4\%$ false positive. It is useful for identifying malicious domains. The agglomerative algorithm shows the best performance. In agglomerative clustering, on average, 94% of the campaign domains are present. Among the 15 clusters, 8 clusters are 100% campaigns domains. The agglomerative algorithm is used here. Paper [44] analyzes the problems like redesigning the algorithms in order to use the distributed computation resources effectively and to come up with a solution for complex hierarchical clustering algorithm CLUBS+ effectively. The high-quality clusters are grouped within their centroids. The accuracy and scalability are achieved using map-reducing algorithms. This is for clustering the big data in a range but still informative version. High accuracy is attained here. Here it uses parallel version of CLUBS+. CLUBS+ helps to combine the benefits of agglomerative and divisive approach. CLUBS+ defines the binary space partition for the domain, leads to the group of cluster that is refined. Agglomerative is a final refinement phase. Outliers are discovered and other points are allocated to the nearest cluster in the refining phase. The clustering quality is improved. It is implemented by using the priority queue. It decreases the number of pairs. It often produces irregular shapes. It improves the quality and identifies the majority of the outlier points. The agglomerative algorithm is used. Paper [45] aims to reduce the human workload, and they are motivated to automate the annotation process. Here the authors consider the online handwriting problems and the Arabic script. They discussed the implementation of word recognition system. Agglomerative clustering is used to produce a Codebook with one stroke per class. In offline handwriting recognition, handwritten documents are described by a scanned image. In online handwriting recognition, handwritten documents are described by a sequence of sample points which represents the writing online data. Their work aims to automate segmentation, labeling, and recognition process. The agglomerative clustering techniques merge at each learning stage, with two closest prototypes to generate a dendrogram. It allows choosing the final number of prototypes at the end of learning.

They select the samples which minimize the sum of distances to the other samples of the same cluster. Here it is combined between two strokes, and relation is composed of five binary values: right, left, above, below, and intersection. The problem in Arabic script is that it is written from right to left. Some characters are represented in shapes and differ from each other by the existence of dots. To avoid segmentation problems, online handwriting is segmented into strokes as written by the writers. The strokes are grouped into a codebook using hierarchical clustering. The agglomerative clustering is used here. In paper [46], opinion mining is used to analyze and cluster the user-generated data like reviews, blogs, comments, and articles. The main objective of opinion mining is to cluster the tweets into positive and negative clusters. In this paper, the authors are able to collect information from social networking sites like Twitter. The meaningful tweets are clustered into two different clusters, positive and negative, using unsupervised ML technique such as spectral clustering. Here the goal is to determine whether the opinion expressed in a twitter is “thumbs-up” or “thumbs-down.” The hierarchical clustering algorithm groups the data objects to form a tree-shaped structure. It is split into agglomerative and divisive clustering. In the agglomerative approach, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. It achieves the best sentiment accuracy. It can solve the problem of domain dependency and reduce the need of annotated training data. It overcomes the problem of clustering multiple files with unlabeled data and performs sentimental classification. In paper [47], the unsupervised feature learning is applied to the mixed-type data to achieve a sparse representation, which makes it easier for clustering algorithms to separate the data. It uses mixed-type data using fuzzy adaptive resonance theory (ART). The ART obtains a better clustering result by removing the differences in treating categorical and numeric features. Benefit of this is demonstrated with several world datasets. It is also demonstrated on noisy, mixed-type petroleum industry data. The goal is to build a framework that automatically handles the differences in numeric and categorical features in a dataset and group them into a similar cluster. It is to reduce the distinction among three numerical and categorical features. It helps to increase the performance. A good cluster has small standard deviation for each cluster as well as averages that vary greatly from one another. The basic clustering model is used here.

58.2.2.3 Divisive

Paper [48] targets to identify the history of arts by grouping the digital paintings based on their features. It automatically learns without any prior knowledge. The performance of clustering is determined by recovering the original groups, F-score, and reliabilities. The spectral clustering algorithm groups the paintings into distinct style groups. The goal of it is finding the characteristics for each paintings by using the N-dimensional feature vectors, and plotting the distance between the endpoints of the vectors. The endpoints are color coded with the low-cost labels. The plots which are scattered are visually inspected for separating the paints. The separation of color provides the semantic information on relation between the styles. The

clustering had an F-score of 0.212. It enables to extract the character of art without any prior information about the nature of the features or the stylistic designation of the paintings. Paper [43] checks a healthcare database which has the patient report for the knee replacement surgery. The patterns of the pain are detected by using the hierarchical clustering. The result indicates the presence of subgroups of the patients based on the pain characteristics. They explain the problem occurring by using the unlabeled medical data. It recognizes the pain-related patterns in the patient report. The national-level dataset is used. Paper [49] analyzes the problems like redesigning the algorithms in order to use the distributed computation resources effectively and to come up with a solution for complex hierarchical clustering algorithm CLUBS+ effectively. The high-quality clusters are grouped within their centroids. The accuracy and scalability are achieved on map-reducing algorithms. This is for clustering the big data in a range but still informative version. High accuracy is attained. Here it uses parallel version of CLUBS+. CLUBS+ helps to combine the benefits of agglomerative and divisive approaches. CLUBS+ defines the binary space partition. It leads to the set of clusters that are refined. The top-down partition is used to separate the hyper rectangle block where the points are close to each other which is equal to minimize the cluster sums of square. It helps to minimize. It splits and replaces the process. It finds the computation of the best split and evaluating the effectiveness of the split. In paper [50], opinion mining is used to analyze and cluster the user-generated data like reviews, blogs, comments, and articles. The main objective of opinion mining is to cluster the tweets into positive and negative clusters. In this paper, the authors are able to collect the information from social networking sites like Twitter. The meaningful tweets are clustered into two different clusters, positive and negative, using unsupervised ML technique such as spectral clustering. Here the goal is to determine whether the opinion expressed in a twitter is “thumbs-up” or “thumbs-down.” The hierarchical clustering algorithm groups the data objects to form a tree-shaped structure. In the divisive approach, all observation starts in one cluster, and splits are performed recursively as one moves down the hierarchy. It achieves the best sentiment accuracy. It can solve the problem of domain dependency and reduce the need of annotated training data. It overcomes the problem of clustering multiple files with unlabeled data and performs sentimental classification.

58.2.2.4 Neural Network

Paper [51] used unlabeled videos from the web to learn visual representations. The idea here is that visual tracking provides the supervision. The authors used auto-encoder for learning the representation based on the ability to reconstruct the input images. It is used to train the unlabeled visual data on the web to train the CNN. Their idea is that two patches connected by a track should have a similar visual representation in deep feature space since they probably belong to the same object. They track millions of patches and learn an embedding using CNN that keeps patches from the same track closer in the embedding space as compared to any

random third patch. The modified version of neural network is used here. In paper [52], sentiment analysis of short texts, such as sentences and Twitter messages, is challenging because of the limited information. To solve it effectively, the authors used some strategies like combining the text content with the prior knowledge. Here they used a deep convolution neural network that exploits from character- to sentence-level information to perform sentiment analysis of short texts. They applied this method for Stanford Sentimental tree bank which has movie reviews and Stanford Twitter Sentimental corpus which contains twitter messages. It achieves 86.4% accuracy. The neural network algorithm is used here. Paper [53] describes the deep learning analysis for sentiment analysis of tweets. Here the authors initialize the parameter weights of CNN, which helps to train an accurate model while avoiding the need to inject any additional features. This method is ranked in the first two positions in both the phrase-level subtask A and on the message-level subtask B. They applied the deep learning model on Semeval2015 Twitter sentiment analysis. The aim of the convolutional layer is to extract patterns. The network initialization process is used to refine the weights of the network passed from unsupervised neural language model. It combines two aspects of IR: unsupervised learning of text representations and learn on weakly supervised data. The basic model of CNN is used here. Paper [54] focuses on supervised technique and uses deep neural network trained on large datasets. Many methods are used such as structured prediction and also multiple pixels are classified by using a network simultaneously. It outperforms the previous algorithm on the area of ROC curve measure. The accuracy is greater than 0.97. The method is also resistant to the phenomenon of central vessel reflex, sensitive in detection of fine vessels (sensitivity >0.87) and fares well on pathological cases. They propose a DL-based method for the problem of detecting blood vessels in fundus imagery, a medical imaging task that has significant diagnostic relevance and was subject to many studies in the past. The proposed approach outperforms previous methods on two major performance indicators, that is, accuracy of classification and area under the ROC curve. Here the authors consider fundus images, that is, pictures of the back of the eye taken in the visible band. A convolutional neural network (CNN) is a composite of multiple elementary processing units, each featuring several weighted inputs and one output, performing convolution of input signals with weights and transforming the outcome with some form of nonlinearity. These results confirm that segmentation of vessels close to the borders of FOVs is more difficult than for the more central pixels. Paper [57] is split into two neural architectures: bidirectional LSTMs and conditional random fields. It constructs and labels the segment by using a transition-based approach and shifts reduce parser. The models depend on two sources of data about the words. The character-based word representations are learned from the supervised corpus and unsupervised word representations are learned from unannotated corpora. The performance in NER in four languages without resorting to any language-specific knowledge or resources such as gazetteers. No language-specific resources or features beyond a small amount of supervised training data and unlabeled corpora are presented by neural architectures for NER. Comparison between the models is made and the scores are reported of the models and with and without the use of

external labeled data such as gazetteers and knowledge bases. The authors obtained an F1 of 91.2 by jointly modeling the NER and entity-linking task. A key aspect of their models is that they model output label dependencies, either via a simple CRF architecture, or by using a transition-based algorithm to explicitly construct and label chunks of the input. Word representations are also crucially important for success; they use both pretrained word representations and “character-based” representations that capture morphological and orthographic information.

58.3 Conclusion

This survey paper gives a clear idea about both supervised and unsupervised learning algorithms. Classification- and regression-based problems can be solved using supervised learning algorithms like support vector machine, KNN, naïve Bayes, and linear regression. Clustering and association mining-based problems can be solved using unsupervised learning algorithms like K-means, divisive, and agglomerative.

References

1. Chikersal P, Poria S, Cambria E. Sentimental analysis of Tweets by combining a rule-based classifier with supervised learning. *International workshop on Semantic Evaluation*, pp 647–651
2. Zemmal N, Azizi N, Dey N, Sellami M (2016) Adaptive semi supervised SVM supervised learning with feature cooperation for breast cancer classification. *Journal* 6:53–62
3. Riaz F, Hassan A, Rehman S, Niazi IK, Dremstrup K (2016) EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. *IEEE Trans* 24(1):28–35
4. Neelakantan A, Collins M (2015) Learning dictionaries for named entity recognition using minimal supervision, 24 April 2015
5. Mansson KNT, Frick A, Boraxbekk C-J, Marquand AF, Williams SCR, Calbring P, Andersson G, Furmark T (2015) Predicting long-term outcome of Internet-delivered cognitive behaviour therapy for social anxiety disorder using fMRI and SVM
6. Devi Arockia Vanitha C, Devaraj D, Venkatesulu M (2015) Gene expression data classification using SVM and mutual information-based gene selection, 13–21, 2015
7. Ozgur Kisi, Kulwinder Singh Parmar (2015) Application of least square SVM and multivariate adaptive regression spline models in long term prediction of river water pollution, 3 August 2015
8. Chen L, Bi L, Si H, Zhang J, Ren Y (2015) Research on prediction method for pivotal indicator of hospital medical quality using decision tree. In: *IEEE international conference*, pp 247–250, 30 November 2015
9. Joshi A, Dangra J, Rawat M (2016) Decision tree based classification technique for accurate heart disease classification and prediction. *Int J* 3(11):1–4
10. Pradhan VM, Vala J, Balani P (2016) A survey on sentimental analysis algorithms for opinion mining. *Int J* 133:7–11

11. Gai K, Qiu M, Elnagdy SA (2016) Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data. In: International conference
12. Yuan Z, Wang C (2016) Online analysis and computing science (ICOACS). IEEE international conference:53–56
13. Sarmah J, Sharma SKR (2016) Decision tree based supervised word sense disambiguation for Assamese. *Int J Comput Appl* 141(1):42–48
14. Jyh-Jian Sheu, Yin-Kai Chen, Ko-Tsung Chu, Jin-Hsin Tang, Wei-Pang Yang. An intelligent three phase spam filtering method based on decision tree data mining
15. Hui LI, Deliang Y, Lingling Y, Yao L, Xia LIN (2016) Supervised massive data analysis for telecommunication customer churn prediction. IEEE international conference
16. Jung YG, Kim KT, Lee B, Youn HY (2016) Enhanced naive bayes classifier for real-time sentiment analysis with SparkR. IEEE International Conference
17. Jiang L, Li C, Wang S, Zhang L (2016) Deep feature weighting for naive Bayes and its application to text classification. *Eng Appl Artif Intell* 52:26–39
18. Nagarajan S, Chandrasekaran RM (2015) Design and implementation of expert clinical system for diagnosing diabetes using data mining techniques. *Indian J Sci Technol* 8(8):771–776
19. Tchakounté F, Hayata F (2017) Supervised learning based detection of malware on Android
20. Kaushik A, Gupta H, Latwal DS (2016) Impact of feature selection and engineering in the classification of handwritten text. IEEE international conference
21. Guinness RE (2015) Beyond where to how: a machine learning approach for sensing mobility contexts using smartphone sensors
22. Dhanalakshmi V, Bino D, Saravanan AM (2016) Opinion mining from student feedback data using supervised learning algorithms. International conference
23. de Silva H, Shehan Perera A (2017) Evolutionary K-nearest neighbour imputation algorithm for gene expression data. In: *Int J*
24. Naghibi SA, Moradi Dashtpajardi M (2017) Evaluation of Supervised learning methods for ground water spring potential mapping in khalkhal region using GIS-based features, pp 169–189
25. Asencio-Cortes, Martinez-Alvarez F, Morales-Esteban A, Reyes J (2016) A Sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction
26. Siwakorn Srisakaokul, Zhengkai Wu, Angello Astroga, Oreoluwa Alebiosu, Tao Xie, “Multiple-Implementation Testing of Supervised Learning Software”, 2018
27. Galan-Garcia P, de la Puerta JG, Gomez CL, Santos I, Bringas PG (2016) Supervised ML for the detection of troll profiles in twitter social network. 24(1):42–53
28. Kusner MJ, Sun Y, Kolkin NI, Weinberg KQ (2015) From word embeddings to document distances. International conference, vol:37
29. Beegum S, Chakraborty D (2016) Identifying cancer biomarkers from leukemia data using feature selection and supervised learning. IEEE international conference
30. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A (2018) Supervised learning of universal sentence representations from natural language inference data, July 2018
31. Rodrigues F, Lourenco M, Ribeiro B, Pereira FC (2017) Learning supervised topic models for classification and regression from crowds. *IEEE Trans Pattern Anal Mach Intell* 39 (12):2409–2422
32. Jiang C, Zhang H, Ren Y, Han Z, Chen K-C, Hanzo L (2017) Machine learning paradigms for next-generation wireless networks. *IEEE Wirel Commun* 24(2):98–105
33. Quanzeng You, Jiebo Luo, Hailin Jin, Jianchao Yang. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia
34. PeirongXu HP, huang T (2018) Unsupervised learning of mixture regression models for longitudinal data. *Comput Stat Data Anal* 125:44–56
35. Li Y, Tao C, Tan Y, Shang K, Tian J (2016) Unsupervised multilayer feature learning for satellite image scene classification. *IEEE Geosci Remote Sens Lett* 13(2):157–161
36. Salamon J, Bello JP (2015) Unsupervised feature learning for urban sound classification

37. Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski V, Konda K, Jean S, Froumenty P, Dauphin Y, Mirza M, Warde-Farley D, Courville A, Vincent P, bangillo Y (2015) EmoNets: multimodal deep learning approaches for emotion recognition on video. *J Multimod User Interface* 10(2):99–111
38. Miotto R, Li L, Kidd BA, Dudle JT (2016) Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 6:1
39. Husic BE, Pande VS (2017) Unsupervised learning of dynamical and molecular similarity using variance minimization. In: Conference
40. Rueda A, Krishnan S. Clustering Parkinson's and age-related voice impairment signal features for unsupervised learning
41. Kulkarni RA (2018) Scrutinizing action performed by user on mobile app through network using ML techniques. *IEEE international conference*
42. Weber M, Wang J, Zhou Y (2018) Unsupervised clustering for identification of malicious domain campaigns
43. Mazzeo GM, Zaniolo C (2016) The parallelization of a complex hierarchical clustering algorithm: faster unsupervised learning on larger datasets
44. Gargouri M, Touj SM, Ben Amara NE (2015) Towards unsupervised learning and graphical representation for on-line handwriting script. *IEEE international conference*
45. Unnisa M, Ameen A, Raziuddin S (2016) Opinion Mining on Twitter data using Unsupervised Learning Technique. Vol:148
46. Lam D, Wei M, Wunsch D (2015) Clustering data of mixed categorical and numeric type with unsupervised feature learning
47. Gultepe E, Edward T (2018) Predicting and grouping digitized paintings by style using unsupervised feature learning. *J Cult Heritage* 31:13–23
48. Khalid S, Judge A, Pinedo-Villanueva R (2018) An unsupervised learning model for pattern recognition in routinely collected healthcare data. In: *International conference*, vol 5, pp 266–273
49. Unnisa M, Ameen A, Raziuddin S (2016) Opinion mining on Twitter data using unsupervised learning technique. *Int J Comput Appl* 148:12–19
50. Wang X, Gupta A (2015) Unsupervised learning of visual representation using videos. *IEEE*
51. dos Santos CN, Gatti M. Deep CNN for sentimental analysis of short texts
52. Severyn A, Moschitti A (2016) Twitter sentimental analysis with deep CNN
53. Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369–2380
54. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition, 17April 2016

Chapter 59

Performance Study of IPv6/IPv4 MANET (64MANET) Architecture



S. Manimozhi and J. Gnana Jayanthi

Abstract Today, the Internet usage has increased due to its efficiency and performance. IPv4 and IPv6 are the most common protocols used for providing Internet communication among the mobile users. There is a big consideration regarding the mobile users' need to access the Internet by different versions of the protocol in MANET. So far, no authors have proposed to design an architecture for intermobility and intertransaction for the transition communication in MANET. For that, 64MANET architecture is proposed to allow the mobile node to roam from one version to another version of the network along with the features of addressing mobility and transition mechanisms of MANET nodes. The performance of 64MANET is evaluated by different performance evaluation metrics. The results show that 64MANET is suitable for intermobility and interoperability between IPv4 mobile node and IPv6 mobile node in MANET communication.

Keywords IPv4 · IPv6 · MANET · Intermobility architecture · Interoperability

Abbreviations

IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
IETF	Internet Engineering Task Force
BGP	Border Gateway Protocol
ISP	Internet Service Provider
NTT	Nippon Telegraph and Telephone
ERNET	Education and Research Network

S. Manimozhi (✉)

Department of Computer Science, Bharathiar University, Coimbatore, Tamil Nadu, India

J. Gnana Jayanthi

PG & Research, Department of Computer Science, Rajah Serfoji Government College, Thanjavur, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_59

AOL	America Online
AT&T	American Telephone and Telegraph Company
MCI	Microwave Communication Commission
VSNL	Videsh Sanchar Nigam Limited
DSTM	Dual Stack Transition Mechanism
DSMIPv6	Dual Stack MIPv6
NAT-PT	Network Address Translation-Protocol Translator
NAPT-PT	Network Address Port Translation-PT
SIIT	Stateless IP/ICMP Translation Algorithm
BIS	Bump in the Stack
BDMS	Bi-directional Mapping System
TRT	Transport Relay Translator
SOCKS	Socket Secure
BIA	Bump-In the-API
MIP-ALG	Mobile Internet Protocol-Application Level Gateway
PET	Prefixing Encapsulation and Translation
DNS	Domain Name System
NAT	Network Address Translation
MAP-T	Mapping of Address and Port using Translation
SIIT-DC	Stateless IP/ICMP Translation for IPv6 Internet Data Center Environments
V4GT	IPv4-Enabled Gateway Translator
V6GT	IPv6-Enabled Gateway Translator
NS	Network Simulator
TCL	Tool Command Language
ByER	Byte-Per-Rate
MTU	Maximum Transmission Unit
PDR	Packet Delivery Ratio
MN	Mobile Node
HN	Home Network

59.1 Introduction

The Internet Protocol version 4 (IPv4) is used for addressing mobile nodes in the network over the past years. The next generation of Internet Protocol (IPv6) developed by Internet Engineering Task Force (IETF) supports the features of address space assignments, security reasons, and interoperability [1]. The IPv6 communication is supported and used by many companies in many countries. The adoption of IPv6 and IPv4 between 2011 and 2018 of network world is depicted in Fig. 59.1.

IPv6 protocol is capable of providing more than 7.9×10^{28} number of addresses than IPv4. In March 2014, 3.4% domains had IPv6 address records in their zones and the entries around 17.4% are found in the Border Gateway Protocol (BGP) routing

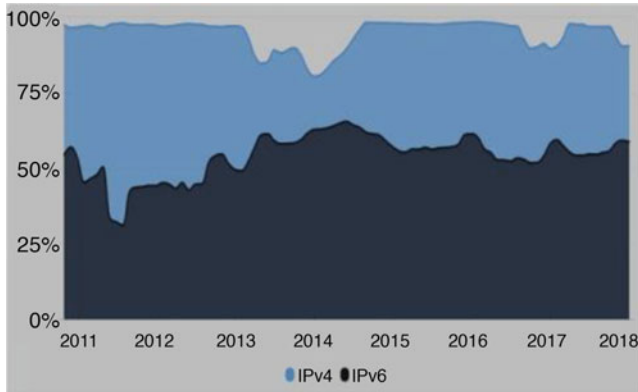


Fig. 59.1 IPv6 adoption in country

table with IPv6 support [2]. Many of the personal computers and server systems are also implemented with IPv6. The first public Internet Service Provider (ISP) of IPv6 was supported by the Telecommunications Company, Nippon Telegraph and Telephone (NTT) in Japan which was announced by itself in March 2000.

In 2008, the Chinese government used the IPv6 infrastructure for summer Olympic Games in the Internet [3]. In 2005, Sify Technologies Limited, a private ISP, India, involutes out IPv6 in 2005, and it was the first to launch a dual-stack commercial. It supports the customers in the way of using dual-stack network for their commercial services. From 2006 onward, the Indian Education and Research Network (ERNET) is providing dual-stack network. Leading enterprises, universities, and government networks have also been deployed IPv6 in their network address.

Since the addresses in IPv4 are running out (220.922/8 address blocks are available for IPv4 Internet), IPv6 is to be deployed in the Internet, ensuring that the Internet can continue to grow and develop. The Internet user's community needs to take their part in the global adoption of IPv6. To motivate and increase the usage of IPv6, the IPv6 Internet Society promoted "World IPv6 Day" on 8 June 2011 [4], and 8 June 2012 was declared as "World IPv6 Launch," as suggested by various website administrators [5]. Realizing the business benefits and the significant features of IPv6 over IPv4, there has been a growth in IPv6 Internet usage [6], as depicted in Fig. 59.2.

IPv6 Forum annotated that, in 2010, IPv6 will become a dominant protocol, and the Internet will become commodity for everyone and everything. The US Government announced that all federal agencies must deploy IPv6 by 2008. IPv6 Forum projected the saturation of Internet to reach 25% by 2010, 35% by 2015, and 50% by 2020 in worldwide. IPv6 infrastructure has been already tested and developed by the network organizations like America Online (AOL) Transit Data Network, American Telephone and Telegraph Company (AT&T), Global Crossing, Microwave Communication Commission (MCI), Nippon Telegraph and Telephone (NTT), Sprint

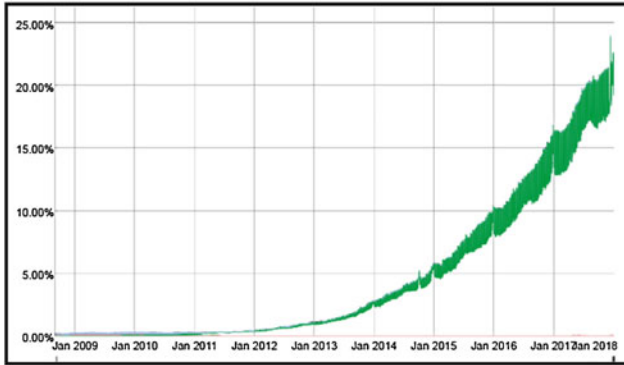


Fig. 59.2 Number of users accessing Google over IPv6

Nextel, Qwest, SAVVIS, Videsh Sanchar Nigam Limited (VSNL)-Teleglobe, Telesonera, France Telecom, and Telefonica [7]. Other than the network organizations, the software development organizations have also used IPv6 in their respective deployments.

The number of Internet users increased day by day based on their needs. In World IPv6 Day meeting, Yahoo pointed out that Yahoo served more than 2.2 million users with IPv6 content for 30 different Yahoo websites [8]. The Facebook served the content to more than 1 million IPv6 users on World IPv6 Day [9]. But this represented only a small fraction – 0.2% – of Facebook users that are IPv6 capable. Of those Facebook users, 0.16% had native IPv6 access and the other 0.04% used 6to4 tunneling.

The IPv4 and IPv6 cannot support intermobility between two different IP networks because the basic functionalities of mobile nodes can support only the communication among the same network and not interoperability. The intermobility architecture which is used to solve the above issue is focused by the following sections: Section 59.2 reviews a study on the mechanisms of transition, Sect. 59.3 provides a concise explanation of the proposed ManIP6/architecture, Sect. 59.4 briefly describes the study of implementation result, and Sect. 59.5 concludes with future direction of research developments.

59.2 Review of Literature Review

There are three transition mechanisms proposed by IETF as (i) dual-stack mechanism, (ii) tunneling mechanism, and (iii) network address translation mechanism. The transition from IPv4 mobile node to IPv6 mobile node is possible in the following cases: (i) IPv6 mobile node is backward compatible with IPv4; (ii) existing IPv4-only equipment (routers, bridges, switches, servers, client nodes, etc.) are to be replaced with IPv4-/IPv6-compatible equipment or IPv6-only

equipment; and (iii) the existing hardware chipset and the existing software version support IPv6. Of all, since IPv6 is not backward compatible with IPv4, there is a need for other new mechanisms. Most of the researchers proposed the architectures by implementing transition mechanisms, and a few of them are Dual-Stack Transition Mechanism (DSTM) [10], Dual-Stack MIPv4 [11], Dual-Stack MIPv6 (DSMIPv6) [12], RoamIP [13], Virtual Overlay [14], 64 Translation as Residential Gateway [15], Network Address Translation-Protocol Translator (NAT-PT) [16], Network Address Port Translation-PT (NAPT-PT) Protocol [17], Stateless IP/ICMP Translation Algorithm (SIIT) [18], Bump in the Stack (BIS) [19], Bi-directional Mapping System (BDMS) [20], Transport Relay Translator (TRT) [21], Socket Secure (SOCKS)64 [22], Bump-In the-API (BIA) [23], Mobile Internet Protocol-Application Level Gateway (MIP-ALG) [24], Prefixing Encapsulation and Translation (PET) [25], Domain Name System (DNS)64 [26], Stateful Network Address Translation (NAT)64 [27], Framework for IPv4/IPv6 Translation [28], 464XLAT [29], Mapping of Address and Port using Translation (MAP-T) [30], Stateless IP/ICMP Translation for IPv6 Internet Data Center Environments (SIIT-DC) [31], IP/ICMP Translation Algorithm [32], and Local-Use IPv4/IPv6 Translation Prefix [33].

There are various transition architectures which have been proposed by many researcher to solve intermobility roam from IPv4 to IPv6 and vice versa. But, no researcher has proposed architecture for IPv4 mobile nodes willing to communicate in IPv6 MANET and IPv6 roaming in IPv4 MANET. The proposed system presents architecture for affording the intermobility also with interoperability for IPv4 MANET and IPv6 MANET.

59.3 Proposed 64MANET Architecture

The proposed new architecture is referred 64MANET which permits the IPv4/IPv6 mobile users to roam freely either into IPv4 MANET or IPv6 MANET and get serviced for global access. In this proposed architecture, let the IPv4 mobile node (B) from IPv4 region or IPv4 MANET while roaming into IPv6 MANET be called as B'. If the IPv4 mobile node needs to communicate, the IPv4-Enabled Gateway Translator (V4GT) is installed automatically to B'. The communication is established by the leader node referred to as A1 and used as an access point to the Internet which provides global connectivity. IPv6 mobile node (A) from either IPv6 network or IPv6 MANET while roaming into IPv4 MANET is called as A'. The IPv6-Enabled Gateway Translator (V6GT) is installed automatically to B1 by the leader node and used as an access point to the Internet which in turn provides global connectivity. Figure 59.3 depicts the proposed architecture.

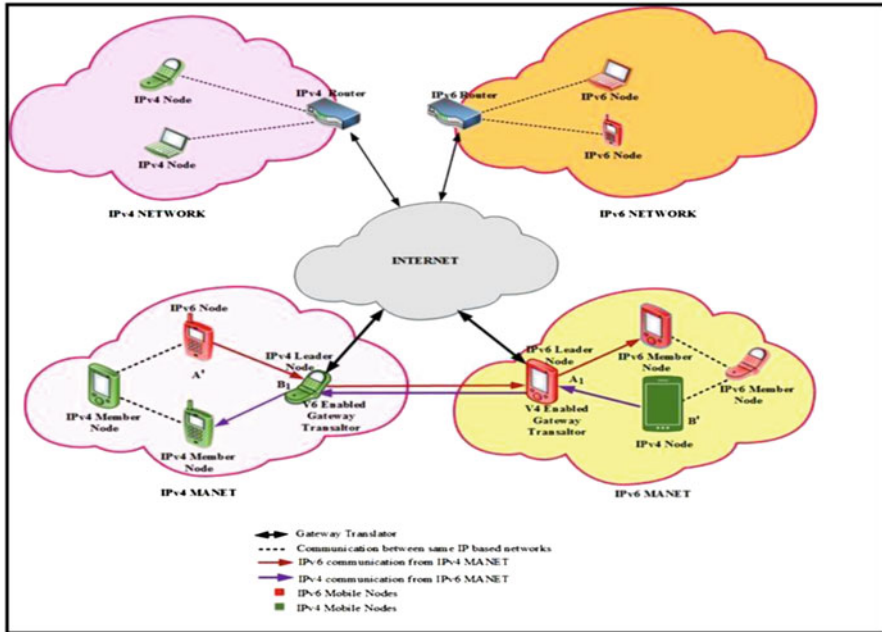


Fig. 59.3 Proposed 64MANET architecture

59.4 Performance Study of 64MANET

59.4.1 Simulation Setup

The designed architecture is simulated using Network Simulator (NS) 2. IPv4 and IPv6 MANET environment are created and the leader node is considered as a router in both IPv4/IPv6 MANET. UNIX Version 16-10-64 operating system is used for simulating the mobile nodes. The dimensions of simulated area are allocated by $1000 \times 1000 \text{ m}^2$, source data rate 20 packets per mobile node, packet size 512 bytes, periodic route update interval 15s, and maximum packets buffered per node per destination 7. Antenna model is omnidirectional, place of nodes is random, and routing model is bidirectional. The version6 Gateway Translator (V6GT) and version4 Gateway Translator (V4GT) interfaces are implemented using Tool Command Language (TCL) script in NS2 simulation languages along with TCP and UDP connections. The proposed simulation is depicted in Figs. 59.4 and 59.5.

The V4GT and V6GT interfaces are installed automatically to the leader node of IPv4 and IPv6, respectively, in order to establish the communication. The important attributes of the simulations are (i) the size of packet considered as in bytes of 512, 1024, . . . 524288; (ii) delay; (iii) IPv4 leader node has the Maximum Transmission Unit as 512 bytes; (iv) IPv6 leader has the Maximum Transmission Unit as 1280 bytes. The important parameters of the study are (1) data loss, (2) payload

Fig. 59.4 IPv4 MANET topology in simulation environment

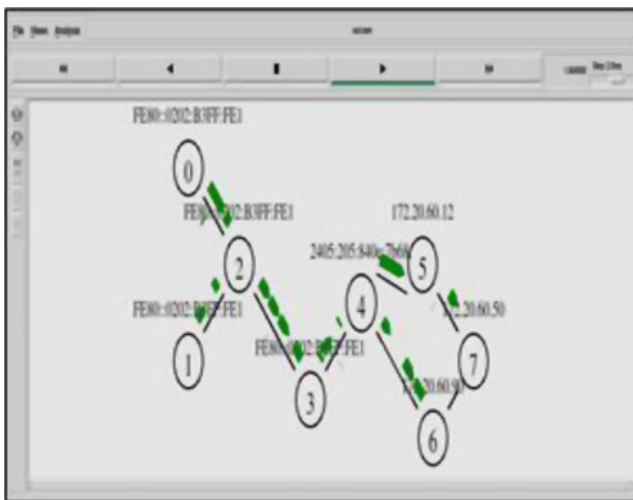


Fig. 59.5 IPv6 MANET topology in simulation environment

overhead, (3) throughput, (4) network latency, (5) end-to-end delay, (6) response time, (7) packet delivery ratio, and (8) handover latency.

59.4.2 Simulation Scenario

The simulation of the proposed architecture is simulated by creating the scenario of (1) IPv4 MANET (5 IPv4 Mobile Nodes and 3 IPv6 MANET nodes with IPv4 leader node act as a router), (2) IPv6 MANET (5 IPv6, mobile nodes, and 3 IPv4 Mobile nodes with IPv6 leader node act as a router), (3) IPv4 mobile node in IPv4 MANET (8 IPv4 MANET nodes), (4) IPv6 mobile node in IPv6 MANET (8 IPv6 MANET nodes), (5) IPv4 MANET only, and (6) IPv6 MANET only within the range of 1000 × 1000 m for the test bed. The test bed has an ability to access the

communications within the boundary region corresponding to access to the Internet. IPv4 MANET created with its same version mobile node along with IPv6 mobile node, and similarly IPv6 network has allowed IPv6 mobile node along with IPv4 mobile node. The leader node functionalities and the gateway functionalities for both IPv4 and IPv6 code in the scripts and also the router coding in the leader node of the MANET. Using this interface, the communication of mobile nodes between two different networks is communicated and roamed.

59.4.3 Simulation Results and Performance Analysis

The main objective of the intermobility simulation is to simulate the IPv4 mobile node roam into IPv6 MANET and IPv6 mobile node roam into IPv4 MANET. At the initial step, the IPv4 leader node initiated the communication with IPv6-intended mobile nodes and IPv6 leader node initiated communication with IPv4 mobile node in MANET. The simulation is based on the interfaces of V4GT and V6GT.

59.4.3.1 Data Loss in Byte-Per-Rate (ByER) Study

This analysis used to study the byte rate ratio between bytes received and transferred bytes. By this analysis, the data loss rate is to be identified between IPv4 mobile node in IPv6 MANET and IPv6 mobile node in IPv4 MANET. The simulated result of data loss analysis is depicted in Fig. 59.6. The data loss rate of the communication between the mobile nodes on the same network has identical rate of data loss and not identical when the source and destined nodes differ in their addressing architecture. The increased data loss during the communication between IPv6 mobile node and IPv4 mobile node is due to the packet size.

Fig. 59.6 Data loss rate comparison

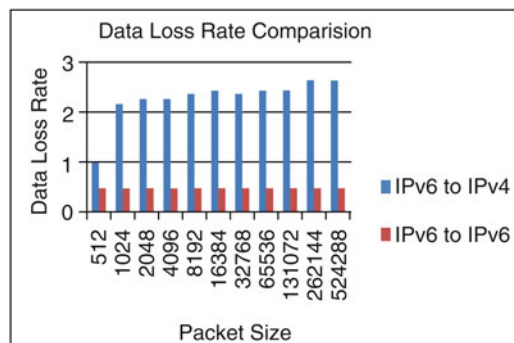
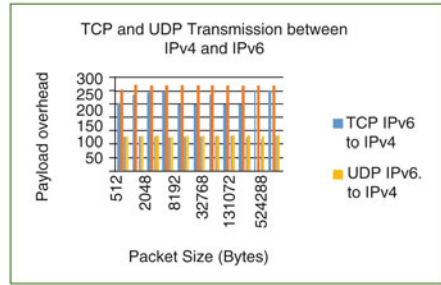


Fig. 59.7 Payload data rate comparison



59.4.3.2 Max Payload Overhead Data Rate Study

The payload data rate increases when the packet size is increased from IPv4 mobile node to IPv6 MANET and IPv6 mobile node to IPv4 in MANET. Both TCP and UDP payload overhead data rates decrease when the communication is established from IPv4 mobile node to IPv6 mobile node than from IPv4 mobile node to IPv4 mobile nodes, from IPv6 mobile node to IPv6 mobile node, and from IPv6 mobile node to IPv4 mobile node in MANET. The study reveals that the payload overhead data rate increases based on fragmentation in each packet of IPv6 mobile node rather than IPv4 mobile node. The simulated results for the payload overhead data rate for UDP transmissions analysis are examined and depicted in Fig. 59.7.

The results of this study imply that if the packet size increases, then the TCP and UDP payload overhead rates also increase. In the simulated results, the TCP and UDP payload data rate increased when the communication is established from IPv4 mobile node to IPv6 MANET. Also payload increased for the communication established from IPv6 mobile node to IPv4 MANET. The TCP and UDP data rates are decreased when the communication is established from IPv6 mobile node to IPv6 mobile node, from IPv6 mobile node to IPv6 MANET, and from IPv6 mobile node to IPv4 MANET. A comparison is also made between TCP payload and UDP payload for the communication established from IPv4 mobile node to IPv6 MANET and from IPv6 mobile node to IPv4 MANET.

59.4.3.3 Throughput Study

The discrepancy of processing overhead and delays in packet size were studied by throughput. The throughput value is increased when the packet size is large. The simulated result shows that the packet transmission from IPv6 mobile node to IPv4 mobile node in MANET reached the best performance. The results of this analysis are depicted in Fig. 59.8.

Fig. 59.8 Throughput comparison

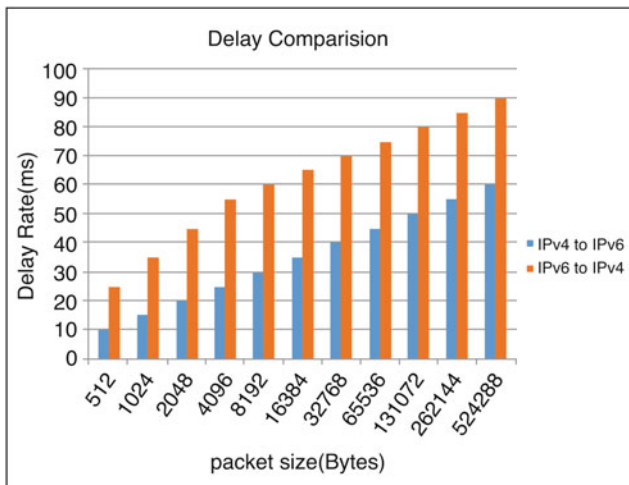
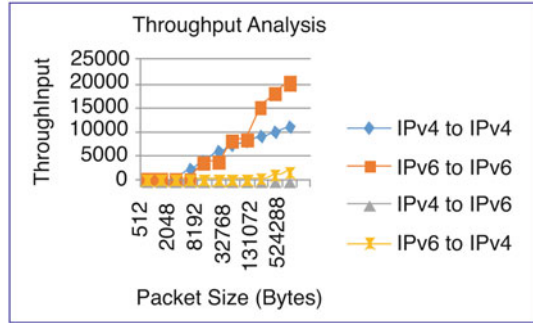


Fig. 59.9 Delay comparison

59.4.3.4 Latency Study

The result of packet transmission from IPv4 to IPv6 MANET reveals less latency, and the transmission from IPv6 mobile node to IPv4 MANET has the highest latency. This study reveals that the IPv6 leader node has too big MTU, and at the receiving end, IPv4 leader node has small MTU. The result is depicted in Fig. 59.9.

The IPv6 leader node receives an ICMPv6 message and retransmits the MTU discover packet with a smaller MTU. This process is repeated until the IPv6 mobile node receives a response that the discover packet arrived whole transmitted packet. For IPv4 communication, the packets were easily transmitted to the IPv6 destined amount of delay for the smaller size packets and larger size IPv4 packets was transmitted with suitable delay.

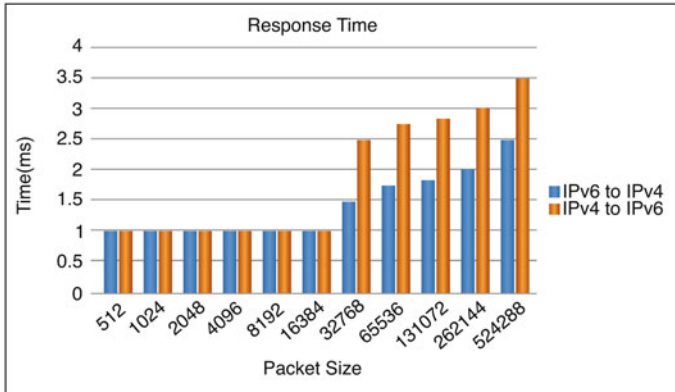
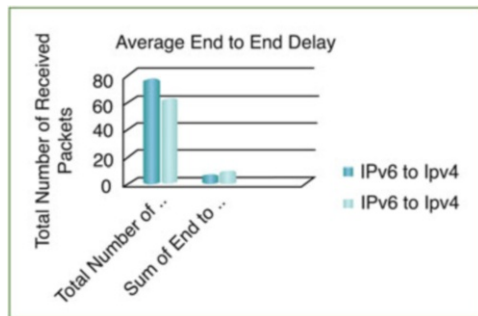


Fig. 59.10 Response time

Fig. 59.11 End-to-end delay



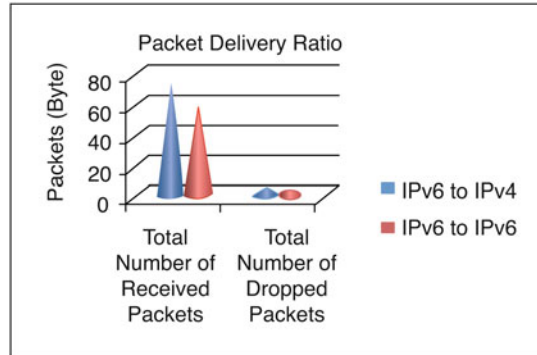
59.4.3.5 Response Time Study

This study analyzed the total time between send requests and receive response between two hosts. In Fig. 59.10, the FTP traffic occurred when 32768 to 524288 KB were received. The simulated result shows that the response time from IPv4 mobile node to IPv6 mobile node and IPv6 mobile node to IPv4 mobile node is low when receiving packets of 512–16384 KB. The response time is increased for IPv4 to IPv6 when compared to IPv6 to IPv4 due to its address translation time.

59.4.3.6 End-to-End Delay Study

The delay will be calculated based on each hop to the destination of each packet. If the delay is lesser, then the packets will be delivered faster to the destination. In the simulation, the number of hops is the same for both networks. Figure 59.11 depicts that the average end-to-end delay of IPv6 to IPv4 is less than IPv4 to IPv6. IPv6 to

Fig. 59.12 Packet delivery ratio



IPv4 has less end-to-end delay as 6.666 s/packet when compared with IPv4 to IPv6 as 11.666 s/packet.

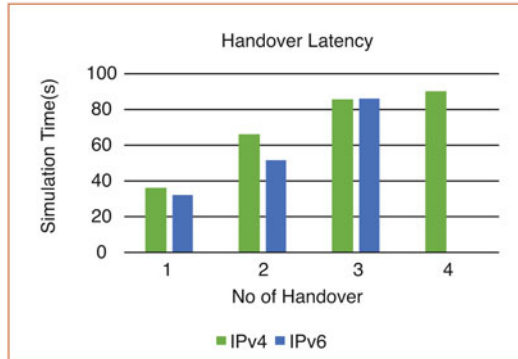
59.4.3.7 Packet Delivery Ratio (PDR) Study

The ratio of successfully received packets to the total sent is referred to as packet delivery ratio (PDR). PDR is applied to IPv4 to IPv6 and IPv6 to IPv4 with FTP and TCP connections. Figure 59.12 shows that when the number of connections increased, the difference in PDR also increased from IPv4 to IPv6 and IPv6 to IPv4 networks. In this simulation, the number of packets is dropped due to their mobility and connectivity of nodes. The PDR depends on its neighbor's response and the speed of nodes. The results show that when the number of connections increases, the PDR decreases. When the PDR rate increased, packet loss decreased in IPv6. The IPv6 has PDR as 94.80% and IPv4 has 93.54%. The handover latency is high due to its triangle routing. In IPv6, the handover latency is low by using the concept of route optimization.

59.4.3.8 Handover Latency Study

Handover latency is one of the most potential factors to evaluate the architecture. It occurs when the IPv4/IPv6 mobile node (MN) moves from its Home Network (HN) to the IPv6/IPv4 MANET, respectively. The time to get an IP address of a new FA registering it into HA is called handover latency. In the simulation time of 100 s for IPv4 mobile node, the reestablishment of connection between MN with HA occurs within 35 s, second reestablishment of connection in 65 s, third reestablishment connection in 85 s, and fourth reconnection in 90 s due to its movement of leader node and MN. For IPv6, the reestablishment of connection between MN and HA occurs within 30 s, second reestablishment of connection in 50 s, and third reestablishment connection in 85 s due to its movement of leader node and MN. Figure 59.13 depicts the handover study between IPv4 and IPv6.

Fig. 59.13 Handover latency



From this study, the IPv4 MN needs higher latency delay when compared to IPv6 MN. The mobile node is communicated in the MANET environment for some particular time and reestablishment needs when the mobile node moves within a boundary frequently. For IPv4, the handover latency is high due to its triangle routing. In IPv6, the handover latency is low by using the concept of route optimization.

59.5 Conclusion and Future Work

The proposed architecture is designed for providing interoperability and intermobility between IPv6 and IPv4 mobile node in MANET. By using 64MANET, the mobile nodes of IPv4 and IPv6 can roam freely in IPv6 and IPv4 MANET without adding any additional hardware requirements. The architecture is tested by NS2 with various packet sizes, delay, and MTU. The V6GT and V4GT interfaces are installed to the leader node which initiates the communications. The proposed architecture can be extended by adding additional security mechanisms for the interoperability between IPv4 and IPv6 mobile nodes when roaming into different versions of the MANET.

References

1. Bradner S, Mankin A (1995) The recommendation for the IP next generation protocol. RFC 1752
2. <http://www.networkworld.com/article/2198933/lan-wan/yahoo-worries-ipv6-upgrade-could-shut-out-1-million-internet-users-initially.html>. Last visited on 2017-12-10
3. <http://www.networkworld.com/article/2179637/lan-wan/yahoo%2D%2Dfacebook-and-google-to-ietf%2D%2Dwhere-are-the-ipv6-users-.html>. Last visited on 2017-12-19
4. <https://www.google.com/intl/en/ipv6/statistics.html>. Retrieved 5.1.2018
5. "The IPv6 Forum". The IPv6 Forum. Retrieved 2017-01-20

6. "Internet Society, Number Resource Organization, and Regional Internet Registries Reinforce Importance of IPv6 Deployment for the Future of the Internet". Internet Society, 2012
7. http://www.alliedtelesis.com/media/fount/white_paper/ipv6_WP_RevD.pdf
8. <http://www.ipv6.ru/english/presscenter/project.php>. Last visited on 2016-3-12
9. ftp://ftp.cordis.europa.eu/pub/ist/docs/rn/ipv6_booklet.pdf
10. Bound J, Toutain L, Richier JL (2005) Dual Stack IPv6 Dominant Transition Mechanism (DSTM), draft-bound-dstm-exp-04.txt
11. Tsirtsis G, Park V, Soliman H (2008) "Dual Stack Mobile IPv4". Internet draft draft-ietf-mip4-dsmipv4-06, IETF
12. Hesham Solima (2008) Mobile IPv6 support for dual stack Hosts and Routers. Internet Draft draft-ietf-mext-nemo-v4traversal-03, IETF
13. Turanyi ZR, Szabo C (2001) Global internet roaming with RoamIP. ACM J Mobile Comput Commun Rev 4(3):58–68
14. Changwen Liu (2004) Support mobile IPv6 in IPv4 domains. In: IEEE proceedings of the international conference on Vehicular Technology (VTC). ISBN:7803-8255-2/04, pp 2700–2704
15. Ssang-Hee Seo In-Yeup Kong (2005) A performance analysis model of PC- based software router supporting IPv6-IPv4 translation for residential gateway. In: Proceedings of the fourth annual ACIS International Conference on Computer and Information Science (ICIS'05) 0-7695- 2296-3/05
16. Aoun C, Davies E (2007) Reasons to move the network address translator – protocol translator. RFC 4966
17. Bangnulo M, Matthews P, van Beijnum I (2008) NAT64/DNS64: network address and protocol translation from IPv6 clients to IPv4 servers. Draft-bagnulo-behave-nat64-00
18. Srisuresh P, Egevang K (2001) Traditional IP Network Address Translator (Traditional NAT). RFC 3022
19. Nordmark E (2000) Stateless IP/ICMP Translation Algorithm (SIIT), RFC 2765
20. Ra'ed AlJa'afreh, John Mellor, Irfan Awan (2008) Implementation of IPv4/IPv6 BDMS translation mechanism. In: IEEE second UKSIM European symposium on computer modeling and simulation. ISBN:978-0-7695-3325-4/08, pp 512–517
21. Hagino J, Yamamoto K (2001) An IPv6-to-IPv4 transport relay translator. RFC 3142
22. Kitamura H (2001) A SOCKS-based IPv6/IPv4 gateway mechanism. RFC 3089
23. Lee S, Shin M-K, Kim Y-J, Nordmark E, Durand A (2002) Dual stack hosts using "Bump-in-the-API" (BIA). RFC3338
24. Choi HH, Cho DH (2003) Mobility management based on mobile IP in mixed IPv4/IPv6 networks. In: IEEE 58th proceedings of VTC 2003- Fall, pp 2048–2052
25. Wu P, Cui Y, Xu M, Wu J, Li X, Metz C, Wang S (2010) PET: prefixing encapsulation and translation for IPv4-IPv6 coexistence. In: The Proceedings of Global Telecommunications Conference (GLOBECOM). ISSN.1930-529X, pp 1–5
26. Bagnulo M, Sullivan A, Matthews P, van Beijnum I (2011) DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers. RFC 6147
27. Bagnulo M, Matthews P, van Beijnum I (2011) Stateful NAT64: network address and protocol translation from IPv6 clients to IPv4 servers. RFC 6146
28. Baker F, Li X, Bao C, Yin K (2011) Framework for IPv4/IPv6 translation. RFC 6144
29. Mawatari M, Kawashima M, Byrne C (2013) 464XLAT: combination of stateful and stateless translation. RFC 6877
30. Li X, Troan O, Matsushima S, Murakami T (2015) Mapping of Address and Port using Translation (MAP-T). RFC 7599
31. Anderson T, Steffann S. Stateless IP/ICMP Translation for IPv6 Internet Data Center Environments (SIIT-DC): dual translation mode. RFC 7756, 016
32. Li X, Bao C, Baker F (2016) IP/ICMP translation algorithm. RFC 6145
33. Meynell K (2017) Local-use IPv4/IPv6 translation prefix. RFC 8215

Chapter 60

Internet of Things: A Technical Perspective Survey



S. Margaret Amala and J. Gnana Jayanthi

Abstract The Internet of Things (IoT), one of the recent emerging technologies in ICT, has a great potential impact on how we live and particularly in every walk of our life. This paper overviews the Internet of Things, highlights the enabling technologies for IoT, outlines the architectural building blocks, and presents the research challenges and issues. The paper is aimed to give an eye opener and to direct the researchers toward IoT in order to understand the concepts and be aware of the burgeoning technology.

Keywords Internet of Things (IoT) · Web of Things · Internet of Nano Things

Abbreviations

IoT	Internet of Things
WoTs	Web of Things
ICT	Information and communication technology
API	Application programming interfaces
SDK	Software development kits
RPL	Routing protocol for low-power and lossy networks
CoAP	Constrained application protocol
RFID	Radio frequency identification
6LoWPAN	IPv6 low-power wireless personal area network
TCP/IP	Transmission Control Protocol/Internet Protocol
WSN	Wireless sensors networks
DODAG	Destination-oriented directed acyclic graph
MQTT	Message queuing telemetry transport
AMQP	Advanced Message Queuing Protocol

S. Margaret Amala (✉) · J. Gnana Jayanthi
PG and Research Department of Computer Science, Rajah Serfoji Government College,
Thanjavur, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_60

60.1 Introduction

IoT has been introduced around since the 1990s. IoT has advanced the way from *anytime, anyplace* connectivity for anyone to *connectivity for anything* through *any network path*, leading to a “smarter world.” IoT can provide highly sophisticated and connectivity of devices, services including several protocols, domains, and applications [1]. Presently the Internet is implemented with both versions of Internet Protocol, IPv4 and IPv6, respectively, to serve a billion of users’ needs. IoT is hastily gaining ground in the development of modern communication world and also has a great impact in every walk of our life [2]. IoT represents a system which consists things in the real world, and sensors attached to or combined to these things, connected to the Internet by way of wired and wireless network structure.

This paper reviews Internet of Things under various titles as in the following sections. Section 60.2 briefs the architectural building blocks of IoT. Section 60.3 briefs the technologies that support for IoT which are quite interesting. Some of the major application domains are presented briefly in Sect. 60.4. This paper is concluded with some of the vital critical issues and challenges in Sect. 60.5, which are to be researched further.

60.2 Architectural Building Blocks

From our literature, we categorize the building blocks components of IoT [1–3, 15], and a few of the most important components are presented in this section.

- *Connectivity and Normalization*: Data collected may be from different domains and similarly protocols used in those domains may also be different to each other. Therefore, a common software interface is required in order to ensure accurate data streaming and interaction with all devices.
- *Device Management*: Components are necessary in order to check if the connected devices are working properly. Device management system also updates the software in the connected network.
- *Database*: Data collected from different sources will have different database management systems with different formats. An amalgam cloud-based database is essential in order to handle data in terms of variety, velocity, veracity, volume.
- *Processing and Action Management*: Smart measures are to be executed by sensing specific data with rule-based and event-action.
- *Analytic*: The techniques of big data analytics are adapted to perform a wide range of complex analysis to extract the most reasonable value out of the IoT data stream.
- *Visualization*: From visualization dashboards, viewing patterns and observing trends can be enabled through line/stacked/pie charts/2D/3D models.

- *Additional Tool:* Developers of Internet of Things may be allowed for prototyping, testing, and marketing use cases. They may also be permitted to create apps for visualizing, managing, and controlling connected devices.
- *External Interfaces:* Integration of third-party systems with IT ecosystem can be done with the help of built-in application programming interfaces (API), software development kits (SDK), and gateways.

60.3 Enabling Technologies

IoT deployment is supported with wireless, mobile, and sensing technologies such as communications networks and protocols, hardware devices, and sensors. Wi-Fi, GPRS, 3G/4G/5G, wireless HART 2010, Zigbee, Bluetooth, etc., are the key denominators as they make more options available to the IoT. Figure 60.1 depicts the architecture of IoT.

- *IPv6/RPL/CoAP Protocols:* With an increasing number of devices connected to the global network, IPv6 is an incontrovertible protocol of choice for the Internet of Things, since IPv6 has a zillion addresses. IPv6 plays a central role in the Internet of Things which is growing fast in the market. IETF have standardized

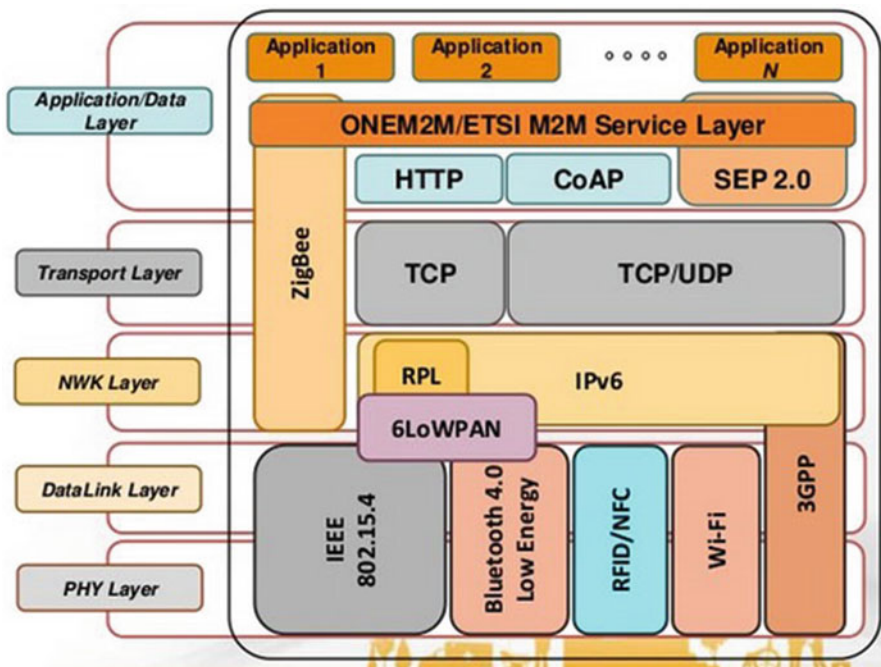


Fig. 60.1 IoT reference architecture

and adopted two new major protocols in IPv6, for the Internet of Things, and they are low-power and lossy networks (RPL) in routing protocol and the **C**onstrained **A**pplication **P**rotocol (CoAP). RPL functions at the IP layer according to the IP architecture and allows for routing across multiple types of link layers. CoAP, the lightweight resources management protocol allows electronics devices to communicate interactively over the Internet [4, 5].

- *Radio Frequency Identification (RFID)*: RFID technology uses electromagnetic fields to read data stored in tags embedded in the items. RFID provides unique identification of objects. RFID system is composed of readers and associated RFID tags which emit the identification location or any other specifics about the object, on getting triggered by the generation of any appropriate signal. The emitted object-related data signals are transmitted to the readers using radio frequencies which are then passed onto the processors to analyze the data [6].
- *6LoWPAN*: Jain et al. (2001) have proposed (IPv6 low-power wireless personal area network) 6LoWPAN architecture by introducing a new adaptation layer between MAC layer and network layer to provide IPv6 support over 802.15.4 running in 2.4 GHz frequency range among 250 kbps transfer rate. This new architecture is shown in Fig. 60.1 and deals with four issues: (i) insert IPv6 packets in 802.15.4 frames, (ii) fragmenting long packets to fit these frames, (iii) stateless compressing packet headers from 48 bytes to 6 bytes, and (iv) forwarding IPv6 packets via multi-hop wireless routes. IPv6 header can be compressed significantly as the nodes of 802.15.4 can commune with their own wireless network. Henceforth, most of the information can be reconstructed from the context of the surrounding 802.15.4 frames which may considerably reduce the number of local addresses that can be used. 6LoWPAN is lightweight and easy to adopt within the small devices [7].
- *Cloud Computing*: The two perspectives of IoT can be classified as (i) Internet centric and (ii) Thing centric. The Internet centric architecture comprises Internet services in which data is provided by the objects. The smart objects take the center stage in the object centric architecture. The full potential of cloud computing is realized if ubiquitous sensing is merged with, so that the sensing service providers can join the network and offer their data using a storage cloud; analytic tool developers can provide their software tools; artificial intelligence experts can provide their data mining and machine learning tools useful for converting information to knowledge, and computer graphics designer can offer a variety of visualization tools. The services provided through cloud computing are classified under three major categories, as (i) infrastructure-oriented services, (ii) platform-oriented services, and (iii) software-oriented services [8]. However, cloud computing brings a major issue in securing the client's data.
- *Nano Technologies*: Nanotechnology reduces the device size in nanometer scale. Nano devices come as sensors and/or actuators [9].
- *Bluetooth*: Bluetooth is a short-range communication protocol, which underlies IEEE 802.15.1, operates in short range of local area network (LAN), and uses only low energy [10].

- *Z-Wave*: Z-Wave is an international standard with 800–900 MHz radio frequency range, developed by Zensys in the year 2001 ranging 100 meters. Z-Wave, a wireless communication protocol, is used for home automation; creates a mesh network using low-energy radio waves to commune between appliances, allowing for wireless control home appliances. It provides interoperability between home control systems of different manufacturers that are a part of its alliance [11].
- *Zigbee*: Zigbee is an international standard with IEEE 802.15.4, developed by Zigbee Alliance, conceived in 1998, standardized in 2003, and revised in 2006, ranging 10–20 meters approximately. Z-Wave, a high-level communication protocol, is used to create personal area networks with small, low-power digital radios, such as for home automation and healthcare automation in a hospital [12].
- *OneM2M*: OneM2M is a global organization founded in 2012. OneM2M specifications offer a framework to support a wide range of applications and services such as smart cities, smart grid, connected car, home automation, public safety, and health [13].
- *Smart-Energy Profile 2.0 (SEP2.0)*: SEP 2.0 connects smart-energy devices in the home to the smart grid. The original work for SEP 2.0 was completed with a joint liaison agreement between the Zigbee Alliance and the Home Plug Alliance which are designed and implemented with Transmission Control Protocol/Internet Protocol (TCP/IP). Even though the standard is still under development, it is expected to be widely adopted by Electric Utilities to implement their consumer facing programs [14].

60.4 Application Domains

Internet of Things is an advanced technological application to multiple consumers and commercial and industrial market segments. The applications for Internet connected devices are extensive. IoT is the next stage of the information revolution and referenced the inter-connectivity of everything from urban transport to medical devices to household appliances [15].

60.4.1 Smart Home and Building

IoT devices are a part of the home automation. Large smart home systems utilize a main hub to provide users with a central control for all of their devices. These devices can include lighting, heating and air conditioning, media and security systems. Wi-Fi has started becoming a part of the home network [16].

60.4.2 Smart Grid

Smart grid requires sensors to provide real-time information. Wireless sensor network (WSN) as smart sensing peripheral information is an important means to promote smart grid technology development and thus promotes the industrial development [17].

60.4.3 Smart City

The smart city includes multimodal sensing information from various IoT platforms into cross-domain and real-time information, using semantic interoperability. Advanced data mining and machine learning techniques are incorporated to provide intelligent actions. The smart city system has to ensure public safety, improve city mobility, and optimize utility usage [18].

60.4.4 Healthcare

Connected healthcare system and smart medical devices monitor personal health and fitness devices, wristband, and other wearable devices. These applications include the potential for remote patient monitoring using smart electronic devices, allowing patients and their doctors to sustain real-time access to health data [19].

60.5 Challenges and Barriers

IoT brings several challenges namely, scalability, mobility management, reliability, availability, interoperability, and security and privacy. In the following, we brief each of these challenges which are the barriers to adoption [2, 18–22].

- *Scalability:* In IoT, things are cooperated within an open environment which includes the basic functionalities, namely, (i) communication and (ii) service discovery functions are required. IoT requires new functions and methods to gain an efficient operation for scalability, more computing power, storage, and battery capacities relatively with low cost and small in size. These trends enable the development of extreme small scale electronic devices with identification, communication, computing capabilities, which could be embedded in other devices, systems, and facilities.
- *Mobility Issues:* IoT devices often move freely and change their Internet Protocol (IPv4/IPv6) addresses, and hence their location becomes dynamic. Routing protocols, such as RPL, have to reconstruct the destination-oriented directed acyclic

graph (DODAG) each time a node goes off the network or joins the network which adds a lot of overhead. In addition, mobility might result in a change of service provider which can add another layer of complexity due to service interruption and changing gateway.

- *Reliability*: System should be perfectly working and delivering all of its specifications correctly. It is a very critical requirement in applications that requires emergency responses. In IoT applications, the system should be highly reliable and fast in collecting data, communicating, and making decisions. However, wrong decisions may lead to disastrous scenarios.
- *Fault Tolerance*: IoT devices are dynamic. Structuring an Internet of Things in a robust and trustworthy manner would require redundancy on several levels and an ability to automatically adapt to changed conditions.
- *Power Supply*: Things typically move around and are not connected to a power supply, so their smartness needs to be powered from a self-sufficient energy source. Although passive RFID transponders do not need their own energy source, their functionality and communications range are very limited.
- *Sensor Networks*: The four most important parameters while integrating wireless sensor networks with Internet are (i) fault tolerance, (ii) energy management, (iii) load balancing, and (iv) security management.
- *Security*: The major problems related to security are (i) authentication and (ii) data integrity. In IoT such approaches are not feasible given that passive RFID tags cannot exchange too many messages with the authentication servers.
- *Standards*: The role of standards is fundamental to make it possible to create open and flexible architectures, and thus boost the diffusion of IoT applications.
- *User-Centric*: A given architecture/application must be useful and easy to use by people who need the service. In other words, a user-centric perspective is necessary in developing new IoT solutions.
- *Cost-Effective*: IoT solutions should be cost-effective. Cost and complexity of software development for the IoT are still a barrier.
- *Business Perspective*: From a business perspective, there is still a lack of compelling success stories to illustrate tangible economic returns and also lack of convincing business models that will be enabled by IoT applications. IoT Developer Survey 2018 outcomes with some noteworthy key findings are as follows:
 1. Amazon AWS and Microsoft Azure are the top two cloud services for IoT. Google Cloud Platform is failing to get traction.
 2. MQTT remains the standard of choice for IoT messaging, while Advanced Message Queuing Protocol (AMQP) is becoming more and more popular as companies scale their IoT deployments and backend systems.
 3. 93% of the databases and data stores used for IoT are open-source software.
 4. Data collected and used in IoT applications is unbelievably miscellaneous.
 5. (v) Data collection and analytics have been identified as one of their top concerns for developing IoT.

From this survey, it is also observed that a very huge amount of data will be collected resulting in big data. Data source has no beginning and no end if IoT is completely implemented. The data streams may vary dynamically and in real scenario may not behave well. This creates a need to focus the research on data stream processing. Further, data ownership presents a critical and ongoing *challenge*, particularly in the social media networks unverified, and introduces nonrelevance and nonaccuracy of data.

60.6 Conclusion

This paper is presented with an overview of IoT. The key findings of the IoT world survey are also presented precisely. We have classified the various challenges and issues in the implementation of IoT which will be more helpful to the researchers interested in IoT. Even though the existing securities proposals provide minimal functionality of security service provisioning, there is a wide room to address these issues.

References

1. Kumar NM, Mallick PK (2018) The Internet of Things: insights into the building blocks, component interactions, and architecture layers. International conference on Computational Intelligence and Data Science. ISBN:1877-0509, pp 109–117
2. Lamaazi H, Benamar N, Jara AJ, Ladid L, El Ouadghiri D (2015) Challenges of the internet of things: IPv6 and network management. In: The proceedings of the eighth international conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS). ISBN:978-1-4799-4331-9, pp 328–333
3. Jayavardhana G, Rajkumar B (2013) Internet of Things (IoTs): a vision, architectural elements, and future directions. J Future Gener Comput Syst Arch 29(7):1645–1660. ISSN:0167-739X
4. Winter T, Thubert P (2012) RPL: IPv6 routing protocol for low-power and Lossy networks. In RFC 6550. ISSN:2070-1721, pp 1–157
5. Shelby Z, Hartke K, Bormann C (2014) Constrained Application Protocol (CoAP). In: Internet draft: intended status: standards track. ISSN:2070-1721, pp 1–112
6. Welbourne E, Battle L (2009) Building the Internet of Things Using RFID: the RFID ecosystem experience. J IEEE Internet Comput 13(3):48–55. ISSN:1089-7801
7. Sadiku MNO, Sarhan M, Musa S, Nelatury R (2016) Internet of Things – an introduction. In: The International Journal of Engineering Research and Advanced Technology (IJERAT). ISSN: 2454-6135, pp 39–43
8. Babu SM, Lakshmi AJ, Rao BT (2015) A study on cloud based Internet of Things: Cloud IoTs. In: The proceedings of the Global Conference on Communication Technologies (GCCT). Electronic ISBN:978-1-4799-8553-1, pp 60–65
9. Akyildiz IF, Brunetti F, Blazquez C (2008) Nano networks: a new communication paradigm. J IEEE Comput Netw 52(12):2260–2279. ISSN:1389-1286
10. Shepherd R (2001) Bluetooth wireless technology in the home. J Electron Commun Eng 13 (5):195–203. ISSN:0954-0695

11. Yassein MB, Wainmardini, AK (2016) Smart homes automation using Z-wave protocol. In: The proceedings of the international conference on Engineering and MIS(ICEMIS), ISBN:978-1-5090-5579-1, pp 1–6
12. <https://www.hindawi.com/journals/wcmc/2018/9857026/>. Last accessed 2018/11/21
13. <http://en.wikipedia.org/wiki/oneM2M>. Last accessed 2018/10/01
14. <http://standards.ieee.org/develop/wg/SEP2.html>. Last accessed 2018/11/10
15. Lee S (2015) An introduction to the Internet of Things from the perspective of the IEEE Internet of Things initiative. In: The proceedings of the international conference on Collaboration Technologies and Systems (CTS). ISBN:978-1-4673-7647-1, pp 506–506
16. Saura S, Torn J (2009) Short communication: Conefor Sensinode 2.2: a software package for quantifying the importance of habitat patches for landscape connectivity. *Environmental Model Software*. ISSN:1364-8152, pp 135–139
17. Song J, Han S, Mok AK, Chen D, Lucas M, Nixon M (2008) Wireless HART: applying wireless technology in real-time industrial process control. In: Proceedings: IEEE symposium on Real-Time and Embedded Technology and Applications (RTAS '08). ISBN:978-0-7695-3146-5, pp 377–386
18. Khera S, Mehla N, Kaur N (2016) Applications and challenges in wireless sensor networks. *Int J Adv Res Comput Commun Eng*. 5(6):448–451. ISSN:2319-5940
19. Islam SMR, Kwak D, Kabir MH, Hossain M, Kwak K (2015) The Internet of Things for health care: a comprehensive survey. *J IEEE* 3:678–708. ISSN:2169-3536
20. Tejasvit K (2014) Challenges in integrating wireless sensor networks into the internet. *Int J Eng Manag Sci* 5(1):7–11. ISSN:2229-600X
21. Chen XW, Lin X (2014) Big data deep learning: challenges and perspectives. *J IEEE Access* 2:514–525. ISSN:2169-3536
22. Zhou Y, Fang Y, Zhang Y (2008) Securing wireless sensor networks: a survey. *Commun Surv Tutor IEEE* 10(3):6–28. ISSN: 1553-877X

Chapter 61

Analysis on DGHV and NTRU Fully Homomorphic Encryption Schemes



B. Santhiya and K. Anitha Kumari

Abstract Homomorphic encryption (HE) is an emerging scheme that allows computation over encrypted data. The standard encryption algorithms like RSA, Elgamal, etc. help in protecting confidential data from attackers rather than performing computation over encrypted data. Fully homomorphic encryption (FHE) permits computation to perform upon encrypted data unlimitedly in server side than in computational node. In this paper, the basic DGHV FHE scheme and NTRU FHE scheme are analyzed to preserve the security and privacy of the data. DGHV performs computing over real integers, while NTRU in a truncated polynomial ring. A detailed investigation of both the schemes is based on the storage and noise reduction that best suits for a real-world application.

Keywords Homomorphic encryption · Fully homomorphic encryption · DGHV · NTRU

Abbreviations

HE	Homomorphic encryption
RSA	Rivest, Shamir, and Adleman
ECC	Elliptic curve cryptography
FHE	Fully homomorphic encryption
DGHV	Diji Gentry Halevi Vaikuntanathan
NTRU	Nth degree truncated polynomial ring units
LWE	Lattice-based encryption
PHE	Partial homomorphic encryption
SHE	Somewhat homomorphic encryption

B. Santhiya (✉) · K. Anitha Kumari
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

LFSR	Linear feedback shift register
PK	Public key
SK	Secret key

61.1 Introduction

To store and to protect the security and privacy of digital data, several methods exist, such as data encryption, tamper-resistant hardwares, etc. When an encryption algorithm does not allow arbitrary computation over encrypted data, the encrypted data must be decrypted before the computation, and the decrypted data is no longer under control. This is where homomorphic cryptosystems can be used in real time to provide better security as these systems enable computations with encrypted data [1].

Homomorphic encryption is the cutting-edge cryptographic technique that allows computations to be carried out in an encrypted data than in plain text to provide privacy and data storage security similar to conventional cryptography, but with added capabilities like searching an encrypted data and performing operations upon encrypted data [2]. At its highest benefit, the service users store and process their data in encrypted format. In addition, encrypted queries are also executed upon encrypted data. Homomorphic encryption can be used in modern -day applications like cloud querying, e-voting, and spam filter [3]. In 1978, Rivest et al. [4] introduced privacy homomorphism. Several algorithms were designed and implemented to improve the accuracy and efficiency of FHE [5]. The DGHV possess both symmetric and asymmetric encryption schemes which are completely based upon the computation over integers [6]. The DGHV scheme's homomorphic behavior is applicable only for a limited circuit depth. Another significant innovation in FHE is NTRU scheme [7]. The operations are based on objects in a truncated polynomial ring (polynomial degree at most $N-1$). Modulus switching and bootstrapping are the two new techniques which support unlimited circuit depth [8].

A detailed review is presented in Sect. 61.2, homomorphic encryption and its properties are elaborated in Sect. 61.3. DGHV scheme is explained in Sect. 61.4. NTRU in Sect. 61.5. The final conclusion is drawn in Sect. 61.6.

61.2 Literature Survey

Khalil Hariss, Maroun Chamoun, and Abed Ellatif Samhat analyzed DGHV and BV-BGV fully homomorphic encryption schemes. The DGHV encryption scheme is based upon the computation over integers. The BV-BGV encryption scheme is based upon the lattice-based encryption (LWE). The main advantage of the

BV-BGV over the DGHV is the modulus switching technique. Bootstrapping can extend the circuit evaluation to unlimited circuit depth. The main drawback is the storage issues, and because of that, these schemes are not practical for real-world applications [9].

Naw Safrin Sattar, Muhammad Abdullah Adnan, and Maimuna Begum Kali proposed a secured aerial photography using homomorphic encryption by considering the problem as the agents to encrypt the images and to upload it on the cloud server that is untrusted [10]. The server performs computation on the encrypted data without the knowledge of the images using NTRU scheme. After concatenation, encrypted result is delivered to the concerned authority for decryption. The main advantage is the use of a server for utilization of recognizing patterns of water, sky, trees, etc. The limitation is that the use of high discretion should be followed for mapping the areas of highly confidential places using images in cloud.

Jiankun Hu and Athanasios V. Vasilakos proposed some methodologies for the problems such as automated and real-time monitoring, automated bill processing, detection of energy loss, early warning of blackouts, fast detection of disturbances in energy supply, real-time energy planning and pricing. These methodologies involve an energy big data-oriented taxonomy for better understanding of the complicated and intriguing relations among various components and its associated solutions. The limitation is that there is a significant reduction in the raw data accumulation and communication [11].

Bingxin Liu and Huapeng Wu proposed a methodology for NTRU that supports somewhat homomorphic encryption (SHE), where a slightly different truncated polynomial ring is used. The methodology used is linear feedback shift register (LFSR) multiplier that can reduce the latency for small input polynomials. The main advantage is that it takes small polynomials with large number of zero coefficients, and thus it significantly reduces the number of cycles. The main drawback is that huge resource is needed to perform these operations [12].

61.3 Homomorphic Encryption

The term homomorphic is an algebraic term homomorphism that refers to mapping between two groups [13]. The different types of homomorphic encryptions are partially homomorphic encryption, somewhat homomorphic encryption, and fully homomorphic encryption [14]. Partial HE allows only one type of operation either addition or multiplication which can be used unlimited number of times. Somewhat HE allows operations like addition and multiplication which can be used only limited number of times. Fully HE allows an unlimited number of operations with unlimited number of times [15].

61.3.1 Homomorphic Scheme

An asymmetric homomorphic encryption scheme ϵ is defined by the following functions,

KeyGen() – It generates a pair of keys termed as public key (pk) and secret key (sk).
 Encrypt (pk, i) – After encryption it outputs a cipher-text ψ_i , where ‘i’ is a plaintext.
 Decrypt (sk, ψ_i) – After decryption it outputs the plain text ψ_i , where ‘ ψ_i ’ is a cipher-text.

Evaluate – Tuple of cipher-texts and public keys are considered as input here. The evaluation function outputs a cipher-text ψ ,

$$\psi = \text{Evaluate}_\epsilon (\text{pk}, C, \psi) \quad (61.1)$$

The scheme is homomorphic if $\psi = \text{Encrypt}_\epsilon (\text{pk}, C, (\pi_1, \pi_2, \pi_3, \dots, \pi_t))$.

61.3.2 Homomorphic Properties

Equation (61.1) states any function “f” which can be represented as Boolean function in a polynomial form. Any polynomial form should be a set of addition and multiplication operations [16]. An FHE scheme must be built in such a way that it satisfies the following homomorphic properties.

61.3.2.1 Addition

$$\text{Enc}_\epsilon (\text{pk}, x_1) + \text{Enc}_\epsilon (\text{pk}, x_2) = \text{Enc}_\epsilon (\text{pk}, x_1 + x_2) \quad (61.2)$$

61.3.2.2 Multiplication

$$\text{Enc}_\epsilon (\text{pk}, x_1) * \text{Enc}_\epsilon (\text{pk}, x_2) = \text{Enc}_\epsilon (\text{pk}, x_1 * x_2), \quad (61.3)$$

where x_1 and x_2 are plain texts in $\{0, 1\}$ and $\text{Enc}_\epsilon (\text{pk}, x)$ is the encryption function.

61.4 DGHV Scheme

DGHV scheme is known as Diji Gentry Halevi Vaikuntanathan scheme. This scheme works only for integers. It can be built upon either symmetrically or asymmetrically [9, 17]. The notations used are listed below:

λ – It is the security parameter used by KeyGen.
 γ – It is the bit length of the integer in the public key.
 η – It is the secret key's bit length.
 ρ – It is the bit length of the noise.
 τ – It is the number of integer in the public key.
 ρ^1 – Secondary noise parameter.

61.4.1 Symmetric Scheme

A secret key is chosen from the interval $p \in [2^{\eta-1}, 2^\eta]$ where η is the bit length of the secret key and the secret key should be an odd integer.

$$C = \text{Encrypt}(p, m) = pq + 2r + m. \quad (61.4)$$

The integers “q” and “r” are randomly chosen, such that $2r$ is smaller than $p/2$.

$$\text{Decrypt}(p, m) = (c \bmod p) \bmod 2. \quad (61.5)$$

As long as the noise “r” is much smaller than “p,” the decryption performs better.

61.4.1.1 Homomorphic Behavior

Assume that we have two plain texts m_1 and m_2 with two cipher-texts $c_1 = pq_1 + 2r_1 + m_1$ and $c_2 = pq_2 + 2r_2 + m_2$, respectively; then the homomorphic properties involves

Addition

$$C_1 + C_2 = p(q_1 + q_2) + 2(r_1 + r_2) + m_1 + m_2 \quad (61.6)$$

Decryption works as long as $2(r_1 + r_2) \in [-p/2, p/2]$.

Multiplication

$$C_1 * C_2 = p (pq_1q_2 + 2r_2q_1 + q_1 m_2 + 2q_2 r_1 + m_1q_2) + 2 (2r_1r_2 + m_2r_1 + m_1r_2) + m_1. \quad (61.7)$$

Decryption works as long as $2 (2r_1r_2 + m_2r_1 + m_1r_2) \in [-p/2, p/2]$. This performs well for limited depths.

61.4.2 Asymmetric Scheme

61.4.2.1 Key Generation

Generate a set of \mathbb{T} public integers as given in the below equation,

$$X_i = p \cdot q_i + r_i \quad (61.8)$$

where $1 < i < \mathbb{T}$ and $X_0 = p \cdot q_0$; “p” is an odd integer to be kept secret.

61.4.2.2 Encryption

1. To encrypt a message, $m \in \{1, 0\}$.
2. A random subset S is chosen from its superset $C \{1, 2, \dots, \mathbb{T}\}$, and a random integer “r” in the range $(-2^{p^1}, 2^{p^1})$.
3. Output of the cipher-text is

$$C = m + 2r + 2 \sum_{i \in S} X_i \text{ mod } X_0, \quad (61.9)$$

where p is kept as secret, q is a random integer, X_i is a public key, and r is a large parameter.

61.4.2.3 Decryption

$$\text{Decrypt } (p, m) = (c \text{ mod } p) \text{ mod } 2 \quad (61.10)$$

61.5 NTRU Scheme

The N th degree truncated polynomial ring units (NTRU) is a public key cryptosystem that uses lattice-based cryptography to encrypt and decrypt the data. NTRU is ahead of other standard encryption algorithms like RSA and elliptic curve cryptography (ECC) as it is resistant to quantum-based attacks. It is not based on the factorization or discrete logarithm problems; rather, it is based on the algebraic structures of certain polynomial rings. Jeffrey Hoffstein, Joseph H. Silverman, and Jill Pipher founded the NTRU in 1996 [10]. The parameters are:

N – The degree of $n-1$ polynomials are present.
 p – each coefficient is reduced by a small modulus (Non-secret).
 q – each coefficient is reduced by a large modulus (Non-secret).
 f – It is the private key polynomial.
 g – To generate the public key h from f this polynomial is used
 h – It is the public key polynomial.
 r – It is the random polynomial (After initial use it is discarded but it is Secret).
 d – Coefficient.

61.5.1 NTRU

Operations are based on objects in a truncated polynomial ring R with degrees $a^0 + a_1 * x + a_2 * x^2 + + a_{N-1} * x^{N-1}$ of at most $N - 1$ is given as,

$$R = \mathbb{Z}[X]/(X^N - 1) \quad (61.11)$$

61.5.2 Key Generation

In the ring R of truncated polynomials, the user B randomly chooses two small polynomials “ f ” and “ g .” The values of these polynomials are secret and must have an inverse of its own. The inverse of f modulo q and the inverse of f modulo p will be computed in such a way that

$$f * f_q^{-1} = 1(\text{modulo } q),$$

$$f * f_q^{-1} = 1(\text{modulo } p). \quad (61.12)$$

Product of polynomials is computed as

$$h = p * ((F_q) * g) \text{ mod } q. \quad (61.13)$$

Private key of B – It comprises a pair of polynomials, namely, f and fp .

Public key of B – It is the polynomial h .

61.5.3 NTRU Encryption

The message “ m ” to be transmitted by the user A will be in the polynomial form. Its corresponding coefficients are chosen from the modulo p between $-p/2$ and $p/2$. To obscure the message, a random number “ r ” is chosen and the encrypted message is

$$e = r * h + m \text{ (modulo } q) \quad (61.14)$$

61.5.4 NTRU Decryption

The message transmitted by user A is received by the user B and is decrypted with the private polynomial “ f ” as

$$a = f * e \text{ (mod } q). \quad (61.15)$$

In an interval of length q , the user B chooses the coefficients of “ A .” The user B computes the original message of A with the private polynomial fp as

$$c = fp * b \text{ (modulo } p). \quad (61.16)$$

61.5.5 Homomorphic Property of NTRU

NTRU possess both the addition and multiplication homomorphic property as shown in the below equation,

$$c_i = h * s_i + pe_i + m_i \in R_q, i \in 1, 2. \quad (61.17)$$

61.6 Conclusion

A detailed pilot study is conducted over FHE schemes based on noise, speed, and key type as the evaluation parameters in this paper. The DGHV is based on computing over real integers, while NTRU is based on objects in a truncated polynomial ring. It is inferred that the DGHV encryption scheme is not the best fit for real-world applications due to time and space overhead, whereas the NTRU utilizes low memory use and generates key at a rapid rate. One main advantage of the NTRU over the DGHV is slower growth of noise than DGHV. Future work sheds light on implementation of NTRU for real-world applications.

References

1. Boneh D, Gentry C, Gorbunov S, Halevi S, Nikolaenko V, Segev G, Vaikuntanathan V, Vinayagamurthy D (2014) Fully key-homomorphic encryption, arithmetics circuit ABE, and compact garbled circuits. Cryptology ePrint Archive. Report 2014/356, 2014. <http://eprint.iacr.org/2014/356>
2. Ramotsoela TD, Hancke GP (2015). Data aggregation using homomorphic encryption in wireless sensor networks. In: 2015 Information Security for South Africa (ISSA), Johannesburg, pp 1–8. <https://doi.org/10.1109/ISSA.2015.7335058>
3. Rivest R, Shamir A, Adelman L (1978) A method for obtaining digital signatures and public-key cryptosystems. *Commun ACM* 21(2):120126
4. Nassar M, Erradi A, Malluhi Q (2015) Pallier's encryption: implementation and clouds applications. *Applied Research in Computer Science and Engineering (ICAR)*, 2015 international conference on, Beirut, pp 1–5. <https://doi.org/10.1109/ARCSE.2015.7338149>
5. Gentry C (2009) A fully homomorphic encryption scheme. PhD thesis, Stanford University, Department of Computer Science
6. Gentry C (2009) Fully homomorphic encryption using ideal lattices. In: *Proceedings of the forty-first annual ACM symposium on Theory of Computing (STOC'09)*. ACM, New York. pp 169–178. <https://doi.org/10.1145/1536414.1536440>
7. Brakerski Z, Vaikuntanathan V (2011) Fully homomorphic encryption from ring-LWE and security for key dependent messages. In: Rogaway P (ed) *Proceedings of the 31st annual conference on Advances in cryptography (CRYPTO'11)*. Springer, Berlin/Heidelberg, pp 505–524
8. Regev O (2005) On lattices, learning with errors, random linear codes, and cryptography. In: *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC '05)*. ACM, New York, pp 84–93. <https://doi.org/10.1145/1060590.1060603>
9. Hariss K, Chamoun M, Samhat AE (2017) On DGHV and BGV fully homomorphic encryption schemes
10. Sattar NS, Adnan MA, Kali MB (2017) Secured aerial photography using homomorphic encryption
11. Hu J, Vasilakos AV (2016) Energy big data analytics and security: challenges and opportunities
12. Liu B, Wu H (2016) Efficient multiplication architecture over ring for NTRU encrypt system
13. Ferrer JD (1996) A new privacy homomorphism and applications. *Inf Process Lett* 60(5):277–282
14. Xiao, Liangliang, Bastani, Osbert and Yen (2012) An efficient homomorphic encryption protocol for multi-user systems. *Citeseer, IACR Cryptology ePrint Archive*, vol 2012, p 193

15. Kipnis, Aviad, Hibshoosh, Eliphaz (2012) Efficient methods for practical fully homomorphic symmetric-key encryption, randomization and verification. IACR Cryptology ePrint Archive, vol 2012, p 637
16. Hariss K, Noura H, Samhat AE (2017) Fully enhanced homomorphic encryption algorithm of MORE approach for real world applications. J Inf Secur Appl 34:233–242. <https://doi.org/10.1016/j.jisa.2017.02.001>. ISSN:2214-2126
17. van Dijk M, Gentry C, Halevi S, Vaikuntanathan V (2010) Fully homomorphic encryption over the integers. EURO-CRYPT (LNCS) 6110:24–43

Chapter 62

Automated Image Captioning for Flickr8K Dataset



K. Anitha Kumari, C. Mouneeshwari, R. B. Udhaya, and R. Jasmitha

Abstract Automated, accurate image captioning is currently a hot topic in the field of deep learning. The model must have the capability to generate human-readable sentences for regions in the image. The model must understand the image to find the words that string together to be comprehensive. To achieve this, in this research work, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are used on Flickr8K dataset. To identify the regions in the image and to recognize the objects in the regions, an advanced region-based CNN (RCNN) methodology has been used. To generate the caption that is most relevant to the image, RNN is used in this paper. Bilingual evaluation understudy (BLEU) score is considered as the evaluation parameter.

Keywords Automated image captioning · Region-based CNN · Recurrent neural network · Flickr8K dataset

Abbreviations

BLEU	Bilingual evaluation understudy
CNN	Convolutional neural networks
NLP	Natural language processing
RCNN	Region-based convolutional neural networks
RNN	Recurrent Neural Networks

K. Anitha Kumari (✉) · C. Mouneeshwari · R. B. Udhaya · R. Jasmitha
PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_62

62.1 Introduction

Humans have the capability to describe a situation immediately upon looking. However, this capability is very difficult for the visual recognition model. Google proposed its own model as Google captioning to create the captions automatically. The limitation of this model is not considering the image region wise to generate the caption. Our model focuses on captioning every region of the image regardless of its position. Deep learning is an emerging technology to develop various products like image recognition software, self-driving car, recommender systems, chatbots, etc. It is very adaptive technology to process various kinds of data. Concepts like batch normalization, dropout, and attention are basic requirements in building a deep learning model. To make the content more accessible and to avoid misunderstanding of the viewer about the image, a noteworthy description about the image is indispensable. Hence, in this research, a detailed, dense description is provided for images rather than a few words. The main challenge involved in this process is to generate the description to an image in the natural language. The other challenge is availability of surplus amount of image datasets where the description must be in context with the location in the image. The descriptions break down into weak words and string them together to form a sentence for any given image is the highest challenge associated.

A ranking-based framework for sentence-based image description and retrieval is established [1]. A holistic data-driven approach to generate description for the images is done by exploiting the vast amount of parallel image data and associated natural language descriptions available on the Web [2]. A system to automatically generate natural language descriptions from images that exploits both statistics gleaned from parsing large quantities of text data and recognition algorithms from computer vision [3]. The training is done with a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes to produce related outputs [4]. Two multimodal neural language models have been introduced to condition on other modalities. An image-text multimodal neural language model can be used to retrieve images when complex sentence queries are fed as input and to generate text conditioned on images [5]. Inspired by recent advances in multimodal learning and machine translation, an encoder-decoder pipeline is also introduced that learns (a) a multimodal joint embedding space with images and text and (b) a novel language model for decoding distributed representations of the space. This pipeline effectively unifies joint image-text embedding models with multimodal neural language models [6]. A model for bidirectional retrieval of images and sentences through a deep, multimodal embedding of visual and natural language data is developed [7]. In order to leverage the information present in all the modalities, one must model the relationships between them effectively. A few techniques have been proposed to tackle this problem by restricting words to visual objects or by

requiring full correspondences between the different modalities [8]. Recent methods for learning vector space representations of words have succeeded in capturing fine-grained semantic and syntactic regularities using vector arithmetic, but the origin of these regularities has remained opaque. The model properties needed for such regularities to emerge in word vectors are analyzed [9]. Semantic propositional content is an important component of human caption evaluation, and a new automated caption evaluation metric defined over scene graphs coined SPICE is proposed [10]. A new deep visual-semantic embedding model trained to identify visual objects using both labeled image data as well as semantic information gleaned from unannotated text is presented [11]. Visual dependency representations have the ability to capture the relationships between the objects in an image, and to hypothesize. This representation can improve image description [12]. A conditional random field model for semantic parsing that reasons jointly about which objects are present in the scene, their spatial extent, and semantic segmentation automatically parses the sentences to extract objects and their relationships and incorporates them into the model [13]. Natural sentential descriptions of RGB-D scenes have been exploited in order to improve the 3D semantic parsing. Importantly, which particular object each noun/pronoun is referring to in the image is reasoned out [14]. Multiple Instance Learning is used to train visual detectors for words that commonly occur in captions, including many different parts of speech such as nouns, verbs, and adjectives. The word detector outputs serve as conditional inputs to a maximum-entropy language model [15]. A system can compute a score by linking an image to a sentence and this score can be used to attach a descriptive sentence to any given image, or to obtain images that illustrate a sentence. The score is obtained by comparing an estimate of meaning obtained from the image to one obtained from the sentence [16].

62.2 Convolutional Neural Networks

A convolutional neural network also known as ConvNet is used to analyze images. It is a deep feed-forward neural network. It is actually influenced by human vision that mimics the vision of human. When a human being sees an object, he does not recognize it on sight. There is a lot of process happening at the back. When a person sees the object in the image, his eyes just do the job of capturing what is around him. It is the brain which does the prior job. First when the eye captures the surrounding, the brain tries to find a pattern of the surrounding based on some correlation between them. These correlations form clusters and the clusters are formed based on color or shape or texture, etc., Finally, the correlations within the clusters are identified and form layers. Thus an object is recognized by forming clusters and then by layers. Computer scientists after knowing this fact thought of mimicking this in technology. And, thus convolutional neural networks emerged.

62.2.1 Working of CNN

Briefing the entire process, the process is similar to sliding a flashlight all over a screen. In the screen the image under consideration and the flash light is the weight matrix and the part of the image that focus on is called the receptive field of the image. Given a large matrix as input to the model, the outcome will be small vectors. Feature extraction and classification are the major processes involved.

62.2.2 Feature Extraction

Dataset Preparation In this step the image is represented in the form of a matrix whose entries are pixels. A three-dimensional matrix is used in this proposed work, as the collected images are colored image (red, blue, and green). The dimensions considered are length, breadth, and height.

Convolution The term convolution refers mixing two different entities and forming a new entity. Thus, in CNN, during convolution step, merge two matrices to obtain the new resultant matrix. In this step sliding of weight matrix is applied over the given image to obtain a new matrix. During this step, irrelevant pixels are denoted with zero and relevant pixels take the numerical values based on their intensity as shown in Fig. 62.1.

Pooling CNN includes pooling layers that are either local or global in which the CNN combines few clusters in one layer into a single neuron and passes to the next layer. There are two types of pooling, namely, max pooling and min pooling. For image captioning max pooling is preferred by fetching the maximum value from each cluster of a layer and passes it to the next layer. The cluster size is a value that is fixed before starting the pooling operation as shown in Fig. 62.2.

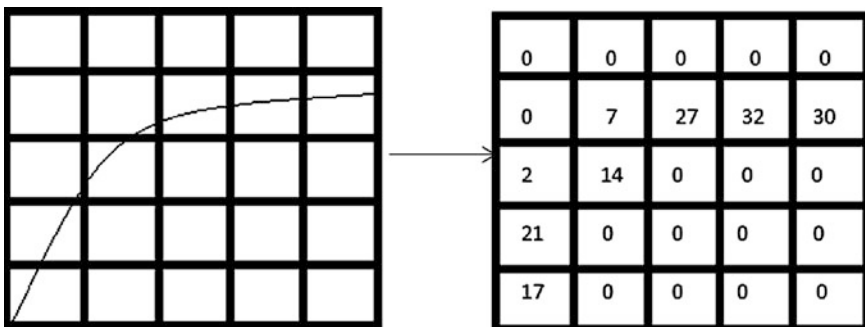


Fig. 62.1 Representation of feature map

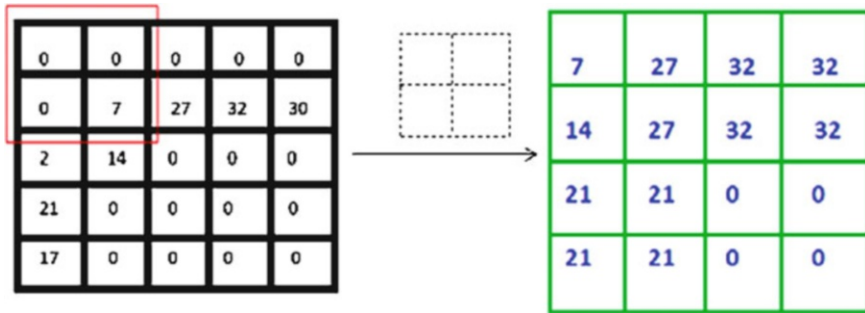


Fig. 62.2 Max-pooling process

62.2.3 Classification

Normalization The distribution range of the features fall in a wide range and the learning rates may lead to the need of correction that differs from one another. Hence, Normalization is preferred and is achieved by Rectified Linear Unit (ReLU).

Regularization A Fully connected layer is prone to over-fitting. It occurs when a model learns the data in depth and too detailed leading to either negative or unexpected outcomes. If it is not prevented, the relations between various regions in an image may lead to an undesired overhead. This can be reduced by dropout. Dropout is a technique used to reduce over-fitting and is achieved by avoiding the complex correlations on data.

62.2.4 Fine-Tuning of Data

It is essential for CNN to be trained with a very large dataset in order to avoid over-fitting. A common way is to enlarge the dataset and to train it from related domains. The biggest advantage of CNN is its accuracy in image recognition. GPU is required to reduce the training time while executing complex tasks. For drug discovery, AtomNet, an extension of ConvNet, is used to predict the chemical features of drugs as ConvNet discovers patterns in an image. CNN also entailed in natural language processing (NLP) operations such as search query retrieval, classification, prediction, and sentence modeling.

62.3 Recurrent Neural Network

Recurrent neural network, precisely called as RNN, is a large funky network with more interconnections between the layers. RNN is preferred in cases where following a sequence is more important. For example, humans are trained to say the alphabets from A to Z effortlessly than from Z to A as they are trained in such a manner. Similarly, in English or in any language, certain rules are to be followed to form a sentence with appropriate words. In such situations, RNN is preferred.

62.3.1 Training of RNN

A set of framed captions is passed as input to NLP parser. The NLP parser accepts the input and analyzes these sentences and produces a graph called “dependency graph” to denote the relationship between the words in the given sentences as shown in Fig. 62.3. Also, it classifies the words in the input sentences as noun, verb, adjective, adverb, conjunction, etc. From the dependency graph, it is possible to determine the possible words in the sentence as a sequence one after the other.

Every layer in an RNN takes two inputs: one as input and other as outcome of the previous layer. While giving the previous layer output to the next layer, normally the outcomes of all the previous layers are given and not just the outcome of the one which is just before it. This is essential for the prediction to be done in a meaningful manner. If simply the word from the previous layer is given, then the dependency graph provides a huge set of words as the next word to be included. This leads to increase the complexity and unwanted iterative predictions. This allows prediction of meaningful sentences with grammatical sequence.

62.4 Region-Based CNN

CNN solely identifies a single object in an image. If the requirement is to identify multiple objects in a single image, region-based CNN is considered as a wise choice. The images are divided into pixels for segmentation that is followed by selective search. Segmentation is dividing the image into pixels to identify whether the



Fig. 62.3 Dependency graph creation

Fig. 62.4 Representation of image in pixels

P1	P2	P3
P4	P5	P6
P7	P8	P9

segments are merged or not. Regions in the image can be identified by a graph-based algorithm. The image is divided into pixels as shown in Fig. 62.4 and the comparison is made based on the intensity values. Here merging the regions is done based on the pixels that have the same intensity values.

62.4.1 Segmentation

This algorithm is based on Felzenszwalb's graph-based segmentation algorithm. The steps are as follows:

- (i) The edges are sorted in ascending order and are marked as $e_1, e_2, e_3,$ and e_4 .
- (ii) Now we have n components to start with the algorithm and we try to merge those components.
- (iii) Repeat for $k = 1, \dots, m$:
 - The segmentation snapshot at the step k is denoted as S_k .
 - We take the k -th edge in the order, $e_k = (v_i, v_j)$.
 - If v_i and v_j belong to the same component, do nothing and thus $S_k = S_{k-1}$.
 - If v_i and v_j belong to two different components C_{k-1i} and C_{k-1j} as in the segmentation S_{k-1} , merge them into one as $w(v_i, v_j) \leq \text{MInt}(C_{k-1i}, C_{k-1j})$; otherwise do nothing.

62.4.2 Selective Search

The regions to be merged are determined by the selective search algorithm. Following are the steps to merge the regions.

- Apply Felzenszwalb and Huttenlocher's graph-based image segmentation algorithm to create regions to start with.
- Similarities between the neighboring nodes are calculated.
- Again the similarities are calculated for another two regions.
- Step 2 is repeated until the image becomes a countable region.

Fig. 62.5 Regions identified in the image



Table 62.1 BLEU score for different algorithms

Model	B-1	B-2	B-3	B-4
Human agreement	61.5	45.2	30.1	22.0
Nearest neighbor	22.9	10.5	0.0	0.0
RNN: Full frame model	14.2	6.0	2.2	0.0
RNN: Region level model	35.2	23.0	16.1	14.8

Let us consider Fig. 62.5 where the image is divided based on the regions. In this figure, the man is divided into many regions. However, it is not preferred as he should be considered as a single region. Thus to address this problem, selective search is used.

These regions and the whole image is given to a pre-trained CNN(VGG-16) algorithm. VGG-16 contains 1024 labels for identifying and mapping into a particular label.

62.4.3 *Generating Captions*

During the training phase, the whole image, along with the regions, is identified and the caption for the corresponding image is generated where the RNN model is pre-trained using the NLP parser. During the testing phase, images are identified as regions and the captions are mapped by RNN. These captions get stringed together to form meaningful captions. Table 62.1 shows the BLEU score evaluations of image region by applying different algorithms.

Although the results are encouraging, the model has few limitations as follows:

- The model can only generate a description of one input array of pixels at a fixed resolution.

- A more sensible approach is required for multiple saccades around the image to identify all entities, their mutual interactions, and wider context before generating a description.
- The RNN receives the image information through additive bias interactions that are known to be less expressive than complicated multiplicative interactions.

62.5 Conclusion

An efficient model is developed in this research work that generates human-readable descriptions for images. Identification of regions is achieved through region-based CNN and the captions are generated using RNN via NLP parser. NLP parser is used to analyze the captions provided to the dataset and is evaluated using the BLEU score. Future avenue is to address the limitations by improving the model effectively.

References

1. Hodosh M, Young P, Hockenmaier J (2013) Framing image description as a ranking task: data, models and evaluation metrics. *J Artif Intell Res* 47:853–899
2. Kuznetsova P, Ordonez V, Berg AC, Berg TL, Choi Y (2012) Collective generation of natural image descriptions
3. Kulkarni G, Premraj V, Dhar S, Li S, Choi Y, Berg AC, Berg TL (2011) Baby talk: understanding and generating simple image descriptions
4. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks
5. Kiros R, Zemel RS, Salakhutdinov R (2014) Multimodal neural language models
6. Kiros R, Salakhutdinov R, Zemel RS (2014) Unifying visual-semantic embeddings with multimodal neural language models
7. Karpathy A, Joulin A, Fei-Fei L (2014) Deep fragment embeddings for bidirectional image sentence mapping
8. Jia Y, Salzman M, Darrell T (2011) Learning cross-modality similarity for multinomial data
9. Jeffrey Pennington R, Manning C (2014) Glove: Global vectors for word representation
10. Chen X, Fang F, Lin TY, Vedantam R, Gupta S, Dollar P, Zitnick CL (2015) Microsoft coco captions: data collection and evaluation server
11. Frome A, Corrado GS, Shlens J, Bengio S, Dean S, Mikolov J et al (2013) Devise: A deep visual-semantic embedding model. In NIPS
12. Elliott D, Keller F (2013) Image description using visual dependency representations
13. Fidler S, Sharma A, Urtasun R (2013) A sentence is worth a thousand pixels. In CVPR
14. Kong C, Lin D, Bansal M, Urtasun R, Fidler S (2014) What are you talking about?
15. Fang H, Gupta S, Iandola F, Srivastava R, Deng L, Dollár P, Gao J, He X, Mitchell M, Platt J et al (2014) From captions to visual concepts and back
16. Farhadi A, Hejrati M, Sadeghi A, Young P, Rashtchian C, Hockenmaier J, Forsyth D (2010) Every picture tells a story: Generating sentences from images. In ECCV

Chapter 63

RAkEL Algorithm and Mahalanobis Distance-Based Intrusion Detection System Against Network Intrusions



R. Padmashani, M. Nivaashini, and R. Vidhyapriya

Abstract With a quick increment in the volume of information in everyday life, there is dependably a requirement for an intrusion detection framework which distinguishes and identifies the attacks at a quicker pace. Any action which abuses the approach of the security premises is characterized as an intrusion. Intrusion detection system (IDS) is an hardware that cocurrences with intrusions produced by an alternate host network frameworks and system sources, then looking at the sign of security issues. IDS is also used to recognize unapproved usage of PC, which ought to interface the gaps in against antivirus and firewall. A general issue in the current IDS is the high false-positives and low-detection rate. This chapter talks about the essential intrusion detection procedures by means of live capturing of network packets. The proposed system uses Mahalanobis distance methodology in best attribute selection and exhaustive search feature selection methods for feature ranking and removal of features for choosing the superlative possible combinations of features from the feature set obtained from the network packets. The RANdom k-labELsets (RAkEL) multi-label ensemble learning algorithm in combination with machine learning algorithms, like J48, support vector machine (SVM) and Naïve Bayes (NB), are utilized to build up the proposed IDS by classifying different network intrusions with higher detection rate and lower false-positive rate.

Keywords Network intrusions · intrusion detection system (IDS) · Wireshark · RAkEL · Mahalanobis distance · J48 · SVM · Naïve Bayes · Weka

R. Padmashani (✉) · R. Vidhyapriya (✉)
PSG College of Technology, Coimbatore, Tamil Nadu, India

M. Nivaashini (✉)
Bannari Amman Institute of Technology, Sathyamangalam, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_63

Abbreviations

ID	Intrusion detection
IDS	Intrusion detection system
PC	Personal computer
RAkEL	RAndom k-labELsets
SVM	Support vector machine
NB	Naïve Bayes
ANN	Artificial neural networks
PCA	Principal component analysis
GA	Genetic algorithm
ARFF	Attribute-relation file format
PCAP	Packet capture
PART	Projective adaptive resonance theory
DoS	Denial of service
U2R	User to root
R2L	Remote to local
TP	True-positive
TN	True-negative
FP	False-positive
FN	False-negative
TPR	True-positive rate
FPR	False-positive rate
MCC	Matthews correlation coefficient

63.1 Introduction

Internet is an infinite worldwide open network. With the growth of the Internet and its potential, there have been subsequent changes in business model of organizations across the world. It is significant to maintain a high-level security to assure security and trusted communication of information between organizations. But secured communication of data over network is under risk of intrusions and attacks. As the Internet emerged into the society, new intrusions or threats like viruses and worms are imported. There are numerous risks of network attacks in the Internet environment, and as a result several systems are designed to block the Internet-based attacks. Particularly, intrusion detection system aids the network to resist external attacks. To overcome this problem, various artificial intelligence and machine-learning methods are developed, such as fuzzy logic, K-nearest neighbor, support vector machine (SVM), artificial neural networks (ANN), Naïve Bayes networks, principal component analysis (PCA), decision tree and genetic algorithm (GA). Among these various methods, an effective method is SVM, but it is well suited for binary classification than for the detection of multiple attacks. Hence, in the proposed IDS, multi-label

ensemble classification algorithm called RAKEL is used along with the binary machine learning classification algorithms. In general, IDS deals with very large amount of data which contain redundant and irrelevant features, by selecting or reducing the features, training and predicting time could be saved. Therefore, in the proposed system two-feature selection methods, best attribute selection and exhaustive search are used by involving Mahalanobis distance calculation measure. The remaining sections of this chapter are organized as follows: Section 63.2 discusses some related work; Sect. 63.3 presents the dataset description and methodology of the proposed system; Sect. 63.4 describes the experimental result and discussion; and Sect. 63.5 provides the conclusions and future scope of the proposed work.

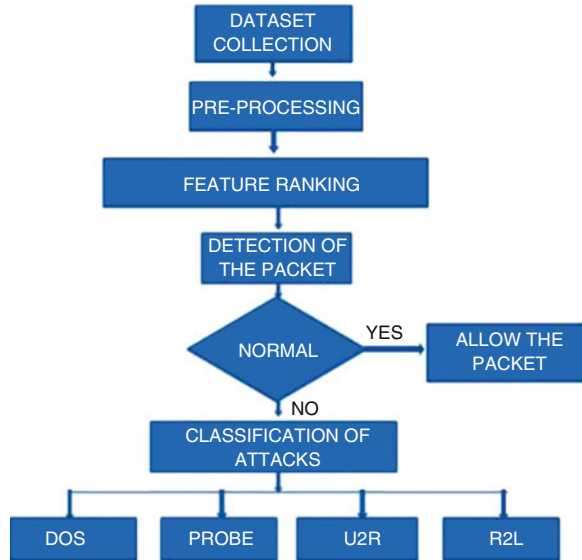
63.2 Related Work

As of late, system security has been the subject of many research works with coming of the Web. There are many works in the composition that analyze about Intrusion Detection System. The genesis of intrusion detection commences with James Anderson's technical report [1], PC Security Threat Monitoring and Surveillance for the U.S. Flying corps. Lee W. Furthermore, Stolfo S., and Mok K. depicted an information digging structure for adaptively building intrusion detection (ID) models [2]. Devale et al. [3] proposed support vector machines (SVM) for twofold grouping, and numerous SVM double classifiers were consolidated with decision trees to have multiclass SVMs, which can arrange the four sorts of assaults. Discriminant investigation is a strategy which can be utilized for picking huge components in huge store of elements [4]. In addition, two characterization frameworks specifically rule-based classification and support vector machine have been utilized for successful grouping of the informational collection [5]. Vitthal Manekar et al. utilize SVM-based IDS for system intrusions with a specific end goal to expand the execution in the exactness of paired characterization used to identify whether the information is malignant or not [6]. Yogita B. Bhavsar et al. utilize SVM technique for arrangement, which can lessen the time required to assemble, demonstrate for grouping and increment the intrusion detection precision [7]. Hence, in the IT industry as well as in the society databases, have gained a great deal. To dissect the helpful data from vast volumes of information, information mining-based IDS has been included.

63.3 Proposed Methodology

The proposed system as shown in Fig. 63.1 makes use of new live captured network packet dataset rather than normal KDD dataset. The features are collected from the packets sniffed using Wireshark. The modules include data collection, pre-processing, and best feature set selection, attack detection, and classification.

Fig. 63.1 Systematic diagram of proposed system



63.3.1 Dataset Collection

The Data collection process is done with a packet sniffing agent called Wireshark [8]. It acts as a monitoring system for IDS in identifying attack and legitimate packets. These data are used in a module for pre-processing the data. Network traffic that is captured includes both malicious and normal packets. The captured data will be saved in a file with .pcap extension. Before pre-processing module, pcap file is exported into .csv format and finally converted into arff format.

63.3.2 Pre-processing and Feature Extraction

Pre-processing techniques are required for information decrease with regard to preparing enormous measure of network traffic with all elements to identify imposters continuously and to give fundamental counteractive action strategies. The nominal data attributes in ARFF file will be converted to numerical attributes and checked for duplication and missing values in the data. Both feature extraction and feature ranking are based on the Mahalanobis distance calculation, calculated among the seven attributes that are listed as follows: packet number, time, source ID, destination ID, protocol, length, and information. All the above-mentioned attributes are obtained from the pre-processing phase and numbered from 0 to 6. For Mahalanobis distance calculation mean and standard deviation is calculated for each attribute in the collected dataset using the following formula

$$D^2 = (X - M)^T C^{-1} (X - M),$$

D – Mahalanobis distance, X – Vector of data, M – Values of free variables, C^{-1} – inverse covariance matrix of free variables, T – Vectors to be transposed.

63.3.3 Feature Ranking and Feature Selection

Feature ranking can be performed either by evaluating individual features (single-feature ranking criteria) or by evaluating subsets of features (subset-based ranking criteria). In feature-ranking process, Mahalanobis distance has been calculated between each individual feature and also between the subset of features in order to find the minimum distance among the features and rank the features based on the minimum distance calculated. Among the seven attributes, destination and information are selected as best attributes initially in the attack detection process.

63.3.4 Rules Generation

By means of the selected attributes, the proposed system uses projective adaptive resonance theory (PART) rule generation algorithm in order to generate rules for the identification of new attacks in the dataset. PART is one of the rule generation algorithms that generate the rules for the normal and abnormal network packet data. Using PART rules generation algorithm, nearly 300 rules were generated for the captured network traffic data. Based on these 300 rules, attacks were detected in the attack detection phase.

63.3.5 Attack Detection and Classification

In the classification process, RAKEL algorithm is used for classifying four major attacks DoS, Probe, User-to-Root (U2R) and Remote-to-Local (R2L) attacks. It is also used with algorithms like J48, SVM, and Naive Bayes (NB) in order to increase the accuracy in the detection rate of the attacks. In the attack detection phase RAKEL algorithm is used along with SVM and Naive Bayes (NB) binary classifiers, where the reduced dataset can be classified as normal and abnormal packets by the binary classifier after that multi-label classification RAKEL algorithm is used on the abnormally labeled dataset, in the same way J48 algorithm is used for decision-making among the reduced dataset for normal and abnormal packet split up, and similarly, as in the former case, RAKEL algorithm can be used on the result produced by the decision-making algorithm.

63.4 Experimental Results and Analysis

The results projected in this section are obtained based on the dataset collected as described in Sect. 63.3. Discovery of normal and abnormal network traffic flow can be estimated by means of resulting measurements:

True-positive (TP): Represents the no. of discovered normal traffic flow, and however it is certainly a normal traffic flow.

False-positive (FP): Represents the no. of discovered normal traffic flow, however it is certainly an abnormal traffic flow.

True-negative (TN): Represents the no. of discovered abnormal traffic flow, and certainly, it is an abnormal traffic flow.

False-negative (FN): Represents the no. of discovered abnormal traffic flow, however, certainly it is a normal traffic flow.

The following are the metrics used as performance measure in the estimation of algorithms like, NB, SVM, J48, and RAKEL.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100 \quad (63.1)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} * 100 \quad (63.2)$$

$$\text{PRECISION} = \frac{\text{TP}}{\text{TP} + \text{FP}} * 100 \quad (63.3)$$

$$\text{RECALL} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100 \quad (63.4)$$

$$\text{F - MEASURE} = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} * 100 \quad (63.5)$$

$$\text{MCC} = \frac{(\text{TP} \times \text{TN}) - (\text{FP} \times \text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} * 100 \quad (63.6)$$

$$\text{ACCURACY} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100 \quad (63.7)$$

From the results as shown in Table 63.1 and Fig. 63.2, it is prominent that the combination of multi-label classifier with the binary classifiers performs well than the individual usage of binary and multi-label classifiers for attack detection and classification. The combination of RAKEL and Naïve Bayes detection model performs better than the other detection models with an accuracy of 96% and low FPR of 2%.

Table 63.1 Analysis of performance metrics for various detection models

Performance metrics	Detection model						
	NB (%)	RAKEL +NB	SVM	RAKEL +SVM	J48	RAKEL +J48	RAKEL
TPR	40.00	61.00	20.00	91.00	40.00	75.00	61.00
FPR	3.00	2.00	4.00	5.00	2.00	2.00	2.00
PRECISION	35.00	92.00	18.00	93.00	35.00	63.00	45.00
RECALL	40.00	95.00	20.00	91.00	40.00	75.00	61.00
F-MEASURE	30.00	93.00	16.00	89.00	36.00	66.00	49.00
MCC	35.00	93.00	14.00	90.00	37.00	67.00	51.00
ACCURACY	55.00	96.00	56.00	95.00	71.00	73.00	60.00

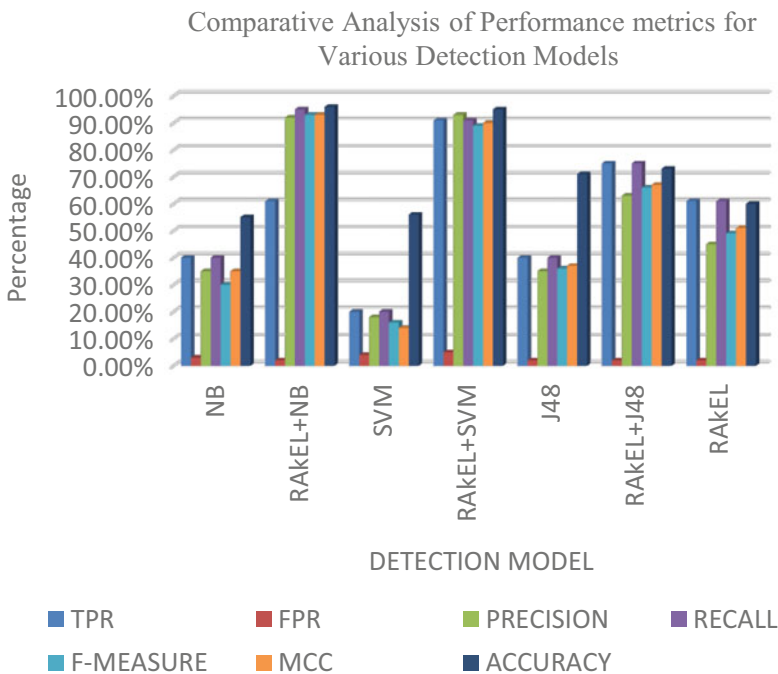


Fig. 63.2 Analysis of performance metrics for various detection models

63.5 Conclusion and Future Work

In the proposed system, Mahalanobis distance approach is used for feature extraction and feature-ranking process in order to reduce the overheads in the correlation methods. This system also proves that machine learning techniques can be used extensively in intrusion detection system for data pre-processing, feature extraction, feature selection, and in detection process. In the proposed system, RAKEL multi-label classification algorithm is combined with the binary classification algorithms

such as SVM and Naive Bayes classifiers. RAKEL algorithm is also combined with decision-making algorithm called J48 in order to increase the performance of the Intrusion Detection System. This system also reduces the false-positive detection rate in the attack detection and attack classification among the data packets captured in the network. Future work is to build Intrusion Prevention System along with the dynamic Intrusion Detection System.

References

1. Anderson JP (1980) Computer security threat monitoring and surveillance. In: Technical report. Fort Washington, Pennsylvania: James P Anderson Co
2. Lee W, Stolfo S, Mok K (2000) A data mining framework for constructing features and models for intrusion detection systems. Technical report, Graduate School of Arts and Science, Columbia University
3. Devale PR, Garje GV, Muly SA (2010) Intrusion detection system using support vector machine and decision tree. *Int J Comput* 3:0975–8887
4. Geetha Ramani R, Siva Sathya S, Sivaselvi K (2011) Discriminant analysis based feature selection in KDD intrusion dataset. *Int J Comput Appl* 31(11):0975–8887
5. Balakrishnan S, Venkatalakshmi K, Kannan A (2014) Intrusion detection system using feature selection and classification technique. *Int J Comput Sci Appl* 3(4):145–151
6. Manekar V, Waghmare K (2014) Intrusion detection system using Support Vector Machine (SVM) and Particle Swarm Optimization (PSO). *Int J Adv Comput Res* 4(3):808
7. Bhavsar Yogita B, Waghmare Kalyani C (2013) Intrusion detection system using data mining technique: support vector machine. *Int J Emerg Technol Adv Eng* 3(3). ISSN:2250-2459
8. Wireshark.org. www.wireshark.org. June 2013. [Online]. Available: <http://www.wireshark.org/>

Chapter 64

Vaguely Node Classification Scheme for Wireless Networks to Design an Intrusion Detection System



S. Latha and V. Sinthu Janita Prakash

Abstract In recent time, wireless networks that are deployed in various applications are facing many problems due to the presence of malicious nodes in the network. In wireless networks, the nodes may exhibit their mal-behavior temporarily or permanently under adverse environmental conditions. Modification of message and dropping the packets are the two main attacks considered in this paper. In these attacks, malicious nodes may behave like a normal node, but involve in unwanted activities. To identify such malicious nodes, a novel vaguely node classification scheme is proposed in this chapter. It may help to derive the rule sets for identifying the malicious nodes in the network which may perform the unknown attack. In turn, it will be more helpful to design an effective Intrusion Detection System for the wireless network.

Keywords Ad hoc on-demand distance vector (AODV) · Dynamic source routing (DSR) · Intrusion detection system (IDS) · Malicious node · Vaguely node · Wireless network

Abbreviations

AODV	Ad hoc on-demand distance vector
DSR	Dynamic source routing
IDS	Intrusion detection system
AHIDS	Advanced hybrid intrusion detection system
MATM	Multi-attribute trust model
DAG	Directed acyclic graph
psure	Positive for sure
nsure	Negative for sure

S. Latha (✉) · V. Sinthu Janita Prakash (✉)
Computer Science, Cauvery College for Women, Tiruchirapalli, India
e-mail: latha.it@cauverycollege.ac.in; sinthujanita.cs@cauverycollege.ac.in

pcount	Positive-count
ncount	Negative-count
DR	Dropping ratio
N_t	Number of transmitted packets
N_r	Number of received packets

64.1 Introduction

In a wireless network, the nodes observe the conditions, detect effects of interest, generate data, and collude in transmitting the data toward a sink. The sink can be a querying user, base station, gateway, or storage node. A network is usually deployed in a hostile and unattended environment to accomplish the data collection and monitoring tasks. When it is used in such type of environment, it faces the problem with malicious nodes and it requires protection. If any compromise occurs in one or multiple nodes, an opponent may involve in several attacks [1] to collapse the network communication.

To deal with many attacks, a broadly utilized countermeasure is multipath forwarding [2, 3], where each packet is forwarded along many routes and hence, the attack may be in some route, but not in all of the routes. This system proposes high extra communication overhead. The other way of countermeasure is to observe the behavior of forwarding nodes [4]. However, these systems may face the problem of high-energy cost. Most of the counter measures are mainly concentrating to identify the attacks when a node alters the messages that too within a certain number of hops. However, without recognizing the intruders, these countermeasures cannot resolve the problems completely because the settled nodes can attack the network continuously without being caught.

Attacks on wireless networks can be categorized as active and passive attacks [5–7], depending on whether the network is intruded or not. Different forms of network attacks are Attacks using Fabrication [8], Resource Consumption Attack [9], Rushing Attack [10], Wormhole Attack [11], Black Hole Attack [12, 13], Jellyfish Attack, Denial of Service [7, 14], Message Tampering, Replay Attack and Replication Attack.

This chapter has six sections. Section 64.2 describes the related work, Sect. 64.3 presents the new classification scheme, simulation environment is given in Sects. 64.4 and 64.5 discusses the simulation results, and the chapter is concluded in Sect. 64.6.

64.2 Related Work

Abirami et al. [15] presented a method to detect the selfish nodes using neighbor credit value-based AODV routing method. Muhammad Saleem Khan et al. [16] identified the misbehaving nodes in MANET with the help of an Adaptive Trust Threshold Strategy. Rupinder Singh et al. [17] presented an advanced hybrid intrusion detection system (AHIDS) which detects the wireless sensor network (WSN)'s attacks automatically. Multilayer Perceptron Neural Network, Feed a Forward Neural Network, Back Propagation Neural Network and Fuzzy Sets are the methods used in this chapter. Ram Prabha et al. [18] implemented a fuzzy logic, multi-attribute trust model (MATM) to check the message success rate, elapsed time of every node, correctness, and fairness in terms of trust metrics. Yang Yu et al. [19] presented many trust factor related to the behaviors of sensor nodes like packet forwarding nature, repetition rate, consistency of the packet content, and the delay etc. Thakurta, Parag Kumar Guha et al. [20] concentrated on wormhole attack in Ad Hoc On-Demand Distance Vector (AODV) protocol. They used round trip bit transfer (RTBT) method to identify the malicious link. Vaseer, Gurveen et al. [21] presented an IDS scheme for probing, DOS, Vampire, and U2R attacks in the mobile ad hoc network (MANET) environment. They use the method of grouping of TCP/UDP and AODV protocols.

64.3 Vaguely Node Classification Scheme

64.3.1 System Model

A typical wireless network with large number of nodes in a two-dimensional area is considered. Each node involves in generation of data periodically, or in forwarding the incoming data packets hop by hop toward a sink (rooting is done using directed acyclic graph (DAG)). The sink may be in any place within the network. It is assumed that all nodes and the sink are time synchronized, which is essential criteria for many purposes. The sink has a complete control over all other nodes and their network topology. It is achieved by any node as it reports about their neighboring nodes soon after it is deployed in the current route. The sink knows the entire information like data, number of packets, the encryption key etc.

It is assumed that the sink is trustful node and it is free from compromising. But the other nodes may be compromised. Compromised nodes may or may not collide with each other.

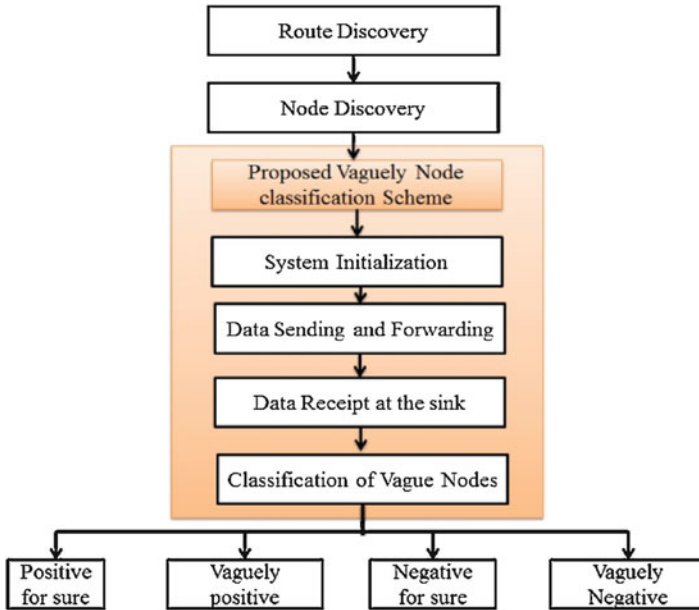


Fig. 64.1 Proposed scheme for the classification of vaguely node

64.3.2 Proposed Scheme for the Classification of Vaguely Nodes

An effective but a simple scheme is presented here to find the intruders in the network by observing the node behavior (Fig. 64.1). As per this scheme, a dynamic routing tree between the sink and the source is first established. The source sends the Hello message toward sink node via a set of intermediate nodes. The sink replies the Hello message through the same intermediate nodes to confirm the route. Then the data are transmitted along the tree structure toward the sink. The source node and the intermediate nodes encrypt the data packets using RC4 scheme and add a small number of extra bits which is called packet marks, to the encrypted data packet. The content of this small packet marks is used to make the sink to get important information of every node. From that packet mark's information, the sink can find out the dropping rate of that particular node. By checking the total number of packets received at the sink, the decryption process and the dropping ratio, the proposed *node classification scheme is used* to identify whether the nodes are surely intruders or vaguely intruder. By changing the tree structure with some specified time interval, behaviors of nodes can be observed in a large variety of different routes. At the end of complete iterations, each node is classified as positive or negative or vaguely positive or vaguely negative. This scheme may be implemented in initial stage of communication or as an intermediate level of transmission.

In this proposed scheme, a system initialization phase is followed by several equal duration rounds of intruders identification phase. It is represented as follows:

Step 1: Route discovery

Step 2: Node discovery

Step 3: Source Node - n^{th} node sends “hello” message to sink via intermediate nodes – Node n, Node n-1, . . . etc. (n is total number of nodes in one iteration/route)

Step 4: Sink receives and returns “Hello” message to the source node (Sink, Node A, Node B, Node n)

Step 5: Working process – source node splits the data packets

Step 5.1: Encrypt the data packets using RC4 Encryption algorithm

Step 5.2: Add some extra bit to the encrypted data packet

Step 5.3: Forward the packet to the next intermediate node

Step 6: An intermediate node receives packets and the intermediate node may

case (i): Drops the packet

case (ii): Forwards packets with step 5 to next intermediate node

case (iii): Modifies the data and then repeats step 5

Step 7: Repeat step 6 until all packets reach sink

Step 8: Finally, sink receives all packets from source via intermediate nodes

Step 9:

Initialize positive for sure(psure)= negative for sure(nsure)=0;

Step 10: For(Nodes $i=1; i<n; i++$)

If ((data(decryption)==data(encryption))

psure[i]= psure+1 **else** nsure[i]= nsure+1

Step 11: Repeat all above for m iterations

Step12: Initialize positive-count(pcount) and Negative-count(ncount) value=0

Step 12.1: Calculating all node’s Dropping Ratio (DR) = $\frac{((N_t - N_r) * N_t)}{(N_t + N_r) + (N_t * (N_t - N_r))}$

N_t - Number of Transmitted Packets & N_r - Number of Received Packets

Step12.2 : for ($i=1; i \leq m; i++$)

if (dropping ratio (DR)==0)

pcount[i] = pcount + 1 **else** ncount[i] = ncount+1

Step 12.3: Node Classification

If ((pcount[i]==m) and (psure== m)) **then** node[i]=Positive

else if ((ncount[i]==m) and (nsure[i]!=0)) **then** node[i]=Negative

else if ((pcount[i]>ncount[i])) **then** node[i]=Vaguely positive

else if ((pcount[i]<ncount[i])) **then** node[i]=Vaguely negative

Table 64.1 NS2 environmental setup

Parameter	Value
Area of simulation	800 m × 800 m
Channel type	Wireless channel
Number of nodes	40,80,120
Data payload	512 B/packet
Traffic type	CBR
MAC types	802_11
Node placement	Random
Transmission range	250 m
Speed	0–20 m/s
Number of malicious nodes	2%, 6%, 10%, and 14%
Pause time	10 s

64.4 Simulation Environment

Table 64.1 represents the simulation setup for this node classification scheme.

64.5 Simulation Results and Discussion

Following figures show the performance of the proposed scheme in detection rate and false-positive rate as simulation parameters in the routing protocols AODV and dynamic source routing (DSR). In this work, AODV and DSR routing protocols are considered, since they work in the dynamic routing environment and both are on-demand routing protocols. Generally, when there are a less number of malicious nodes, the identification of them is much easier. However, after multiple rounds of iterations, the identification, that is, the detection rates under the different percentage of malicious nodes are the same and the scheme produces a good and very high detection rate.

Figures 64.2, 64.3, 64.4 and 64.5 depict the detection rate and Figs. 64.6, 64.7, 64.8 and 64.9 present the false-positive rate of the proposed node classification scheme with 2%, 6%, 10%, and 14% malicious nodes against the number of iterations

64.6 Conclusion

Using this vaguely node classification scheme, the behavior of every node in the network is analyzed. Checking is made on whether the node involves in modification of message and/or dropping of packets without forwarding them. Based on their behavior, every node is categorized as either as positive or negative or vaguely

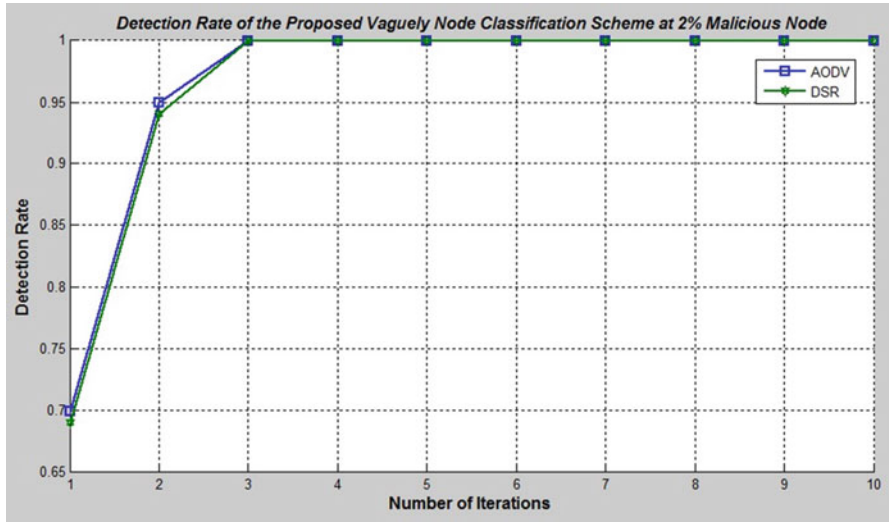


Fig. 64.2 Detection rate – 2% malicious nodes

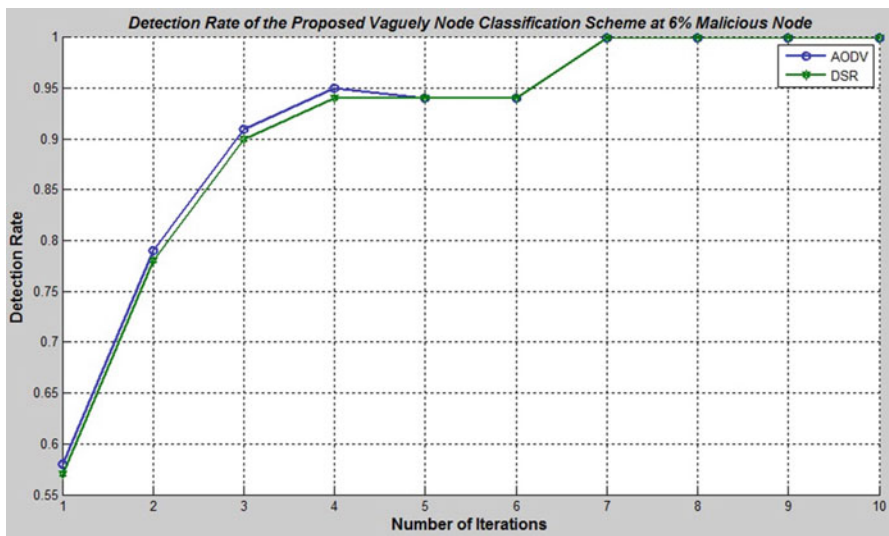


Fig. 64.3 Detection rate – 6% malicious nodes

positive or vaguely negative. The proposed scheme performance is analyzed regarding the detection rate and the false-positive rate when this scheme is implemented in the routing protocols like AODV and DSR. Generally, when there are a less number of malicious nodes, the identification of them is much easier. However, after multiple rounds of iterations, the detection rate under the different percentage of malicious nodes is same and the scheme produces a good detection rate. And both

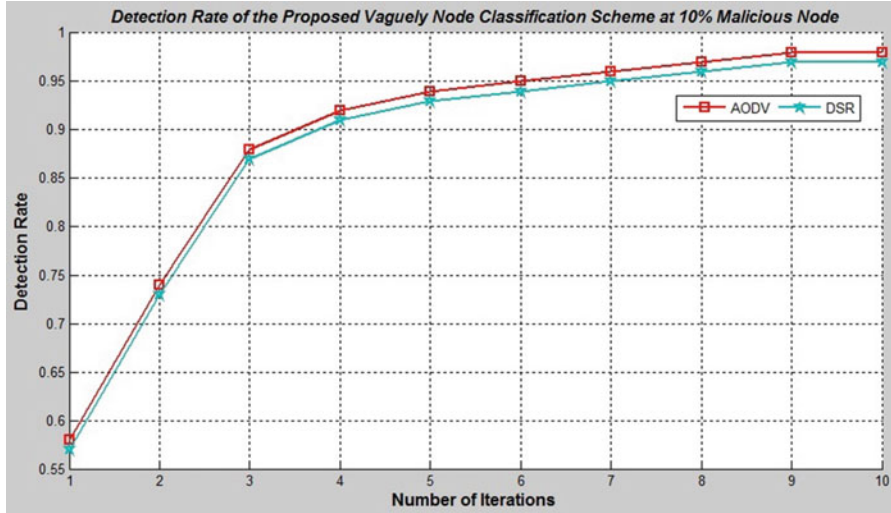


Fig. 64.4 Detection rate – 10% malicious nodes

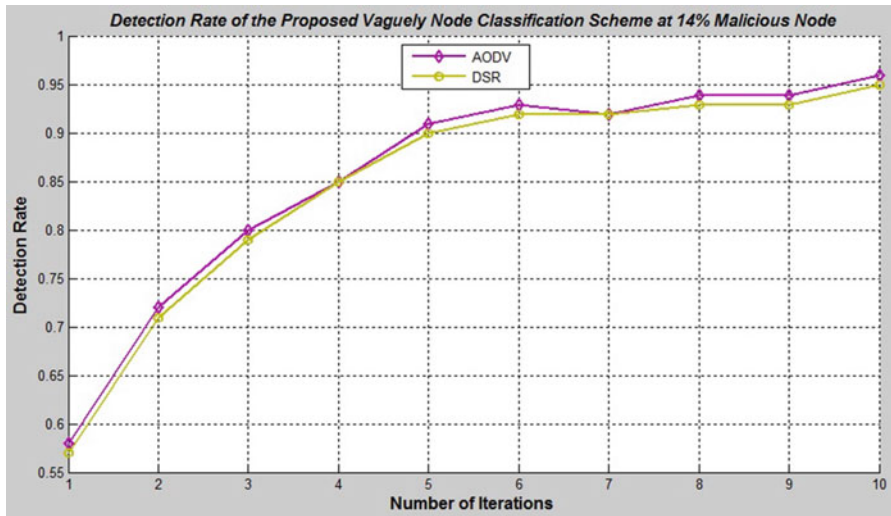


Fig. 64.5 Detection rate – 14% malicious nodes

the algorithms produce acceptable range of the false-positive error rates. From the above results, it is clearly indicated that the proposed classification scheme performs well on both the routing protocols. It is clear that it produces better performance in AODV than DSR. As a future work, unknown attack rule structures can be generated when using this which will be more helpful for designing an effective Intrusion Detection System.

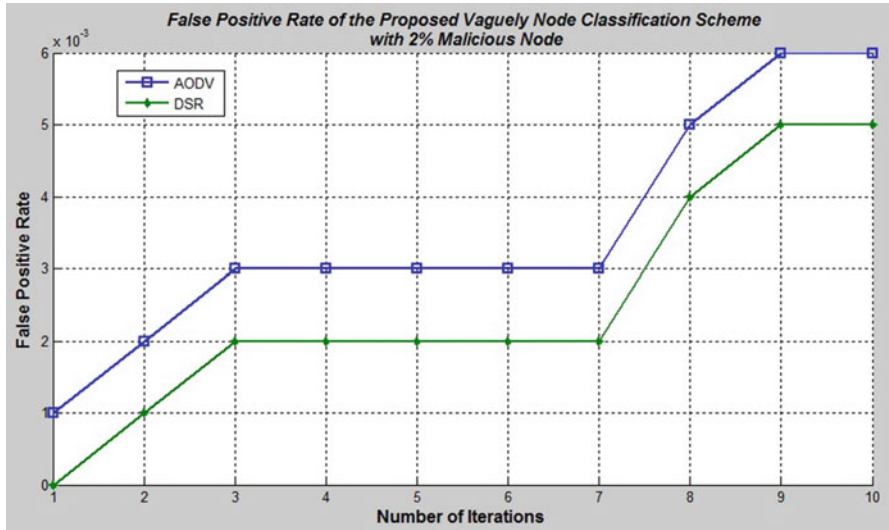


Fig. 64.6 False-positive rate – 2% malicious nodes

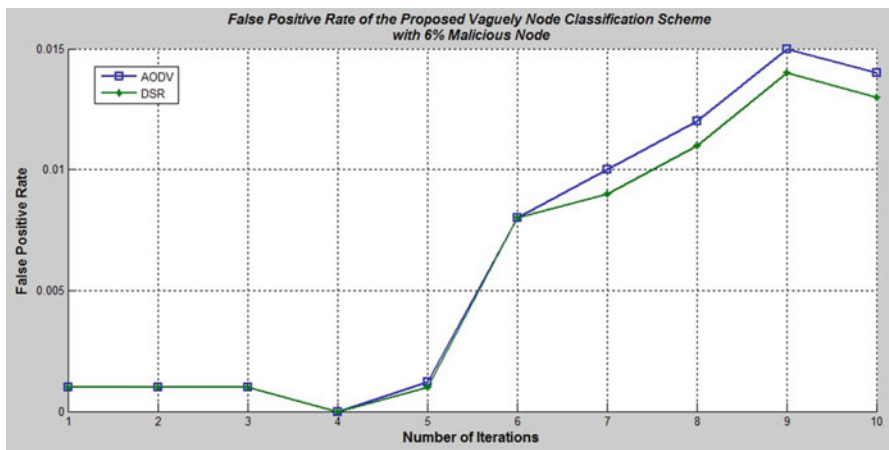


Fig. 64.7 False-positive rate – 6% malicious nodes

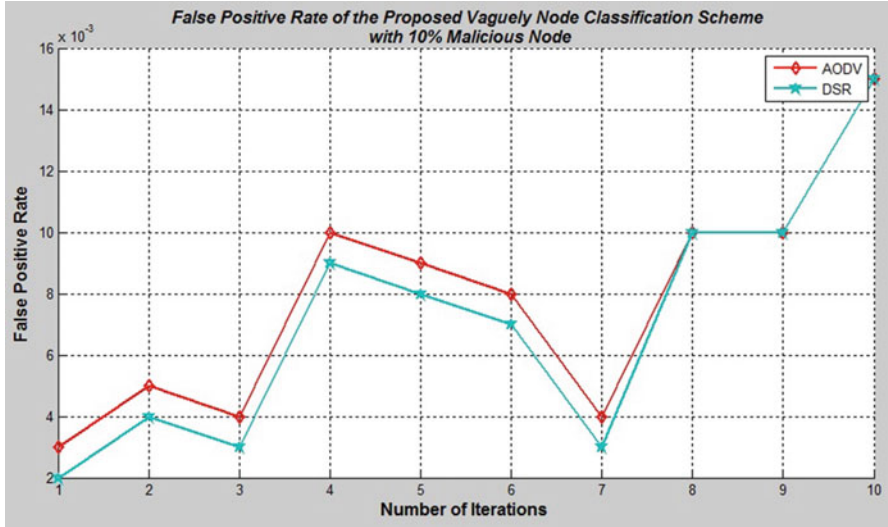


Fig. 64.8 False-positive rate – 10% malicious nodes

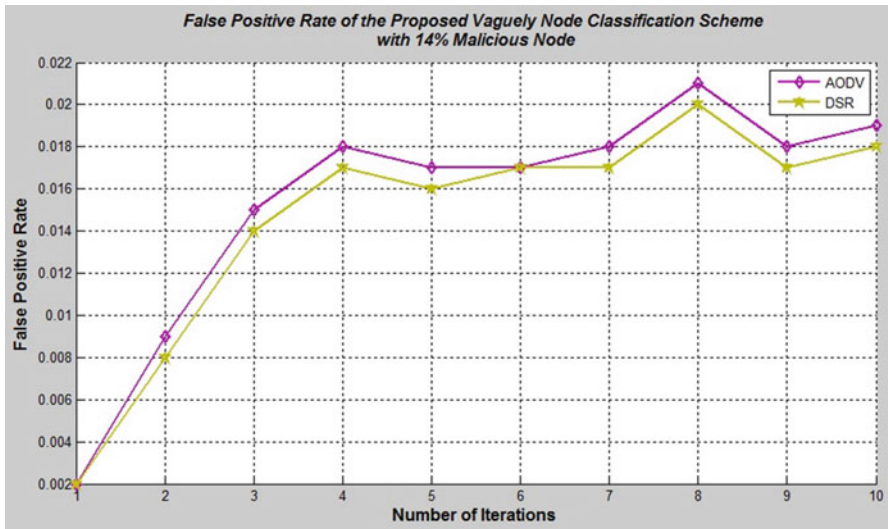


Fig. 64.9 False-positive rate – 14% malicious nodes

References

1. Bhushan B, Sahoo G (2018) Recent advances in attacks, technical challenges, vulnerabilities and their countermeasures in wireless sensor networks. *Wirel Pers Commun* 98(2):2037–2077
2. Devi VS, Hegde NP (2018) Multipath security aware routing protocol for MANET based on trust enhanced cluster mechanism for lossless multimedia data transfer. *Wirel Pers Commun* 100(3):923–940
3. Bouabdellah M et al (2018) Network layer attacks and countermeasures in cognitive radio networks: a survey. *J Inf Secur Appl* 38:40–49
4. Abirami KR, Sumithra MG (2018) Evaluation of neighbor credit value based AODV routing schemes for selfish node behavior detection. *Clust Comput* 43:1–10
5. Latha S, Prakash SJ (2017) A survey on network attacks and intrusion detection systems. *Advanced Computing and Communication Systems (ICACCS)*, 2017, IEEE, pp 1–7
6. Fleming T, Hjalmar W (2018) Network intrusion and detection: an evaluation of SNORT, pp 1–28
7. Osanaiye et al (2018) Denial of service defence for resource availability in wireless sensor networks. *IEEE Access* 6:6975–7004
8. Juneja et al (2018) Security against vampire attack in ADHOC wireless sensor network: detection and prevention techniques. In: *International conference on wireless intelligent and distributed environment for communication*. Springer, Cham, pp 25–38
9. Jain S, Agrawal K (2017) The impact of resource consumption attack on signal-stability based adaptive routing protocol in Manet. *Indian J Sci Technol* 10(30):1–5
10. Borkar GM, Mahajan AR (2017) A secure and trust based on-demand multipath routing scheme for self-organized mobile ad-hoc networks. *Wirel Netw* 23(8):2455–2472
11. Butle P et al (2018) Securing on demand source routing protocol in mobile Ad-Hoc networks by wormhole attacks. *Int Res J Eng Technol (IRJET)* 5(3):2767–2772
12. Singh et al (2018) An improved mechanism to prevent Blackhole attack in MANET. In: *Progress in advanced computing and intelligent engineering*. Springer, Singapore, pp 511–520
13. Sen B et al (2018) A trust-based intrusion detection system for mitigating Blackhole attacks in MANET. *Adv Comput Commun Paradig* 706:765–775. Springer, Singapore, AISC
14. Sicari S et al (2018) REATO: REActing TO denial of service attacks in the internet of things. *Comput Netw* 137:37–48
15. Abirami KR, Sumithra MG (2018) Evaluation of neighbor credit value based AODV routing schemes for selfish node behavior detection. *Clust Comput*, pp 1–10
16. Khan MS et al (2017) Isolating misbehaving nodes in MANETs with an adaptive trust threshold strategy. *Mobile Netw Appl* 22(3):493–509
17. Singh R, Singh J, Singh R (2017) Fuzzy based advanced hybrid intrusion detection system to detect malicious nodes in wireless sensor networks. *Wirel Commun Mobile Comput* 2017:1–14
18. Prabha VR, Latha P (2017) Fuzzy trust protocol for malicious node detection in wireless sensor networks. *Wirel Pers Commun* 94(4):2549–2559
19. Yu Y et al (2017) An efficient trust evaluation scheme for node behavior detection in the IoT. *Wirel Pers Commun* 93(2):571–587
20. Thakurta et al (2018) An efficient approach for detecting wormhole attacks in AODV routing protocol. *Adv Comput Commun Paradig* 706:217–227. Springer, Singapore, AISC
21. Vaseer G, Ghai G, Patheja PS (2017) A novel intrusion detection scheme: an AODV routing protocol case study. *Nanoelectronic and Information Systems (iNIS)*, 2017 IEEE international symposium

Chapter 65

Dynamic Traffic Light Scheduling for Emergency Vehicles Using Fog Computing



S. Sarathambekai, T. Vairam, and A. Dharani

Abstract This chapter presents a dynamic traffic light scheduling system, which schedules the time of the traffic light controller on each side. Each emergency vehicle is assigned with priority and it has a unique identity (ID) by default. The unique ID and the priority are used to process the request from each vehicle. In the case of priority, conflict distance is considered for scheduling. Various technologies like ZigBee, LoRaWAN are used to provide cost-effective solutions. ZigBee is one of the high-level communication protocol in IEEE 802.15.4. It works at low power, and it has low data rate. ZigBee has a range of 10–100 meters with line of sight communication. The rate of data transmission in ZigBee is 250 kb/s.

Keywords Fog computing · Line of sight · LoRaWAN · Transceiver

Abbreviations

LoRaWAN	Long range wide area network
RFID	Radio frequency identification
VIN	Vehicle identification number
USN	Ultrasonic sensor node
GPS	Global positioning system
ETA	Estimated time to arrive
MAC	Media access control
UHF	Ultra high frequency
PLC	Programmable logic controller
MIRT	Mobile infrared transmitter
UIN	Unique identification number
GSM	Global system for mobile communications

S. Sarathambekai (✉) · T. Vairam · A. Dharani
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu,
India

65.1 Introduction

India is a fast growing economy with second largest population in the world. There are congestion problems in the cities. There is a drastic growth in the number of vehicles in the past few years. India requires a smart solution for solving the issue of traffic congestion. Especially, in the case of emergency vehicles, the impact of traffic congestion is loss of life. So, society is in need of a dynamic system which schedules the traffic light at the poles based the type of emergency vehicles and distance at the ZigBee transmits the data. Negative impact of congestion can be reduced by proper intelligent management of traffic flows.

The fog computing is an extension of cloud. The devices involved in fog computing are called as called fog nodes. Fog nodes can be deployed anywhere with a network connection without any specific environment like cloud servers. In real-time routers, switches and embedded servers are called fog nodes. There are various benefits of fog computing namely greater business agility, better security, deeper insights with privacy control, lower operating expenses.

The whole chapter is grouped into five parts. Section 65.2 talks about the literature survey. Section 65.3 talks of how the proposed model will overcome the problems faced in developing countries as well as developed countries for traffic congestion of emergency vehicles. Section 65.4 explains the work in proposed model. Section 65.5 gives the implementation details of the project.

65.2 Related Works

A detailed survey is done about the various techniques used in traffic light controllers for emergency vehicles.

Rajeshwari Sundar et al. [1] developed the prototype of intelligent traffic control system and tested it in wireless communication laboratory. This model consists of three parts. First part contains automatic signal control system. In this system, each vehicle is assigned a RIFD tag. When the vehicle is in the range of radio frequency identification (RFID) reader, it will send the signal to the RFID reader. The RFID reader will track the number of vehicles that have passed through for a specific period and find the volume of congestion. The second part explains the work of emergency vehicle clearance. A ZigBee transmitter is placed at each emergency vehicle and the ZigBee receiver will be placed at the traffic light controller side. A signal will be sent through the ZigBee transmitter to the ZigBee receiver. This signal is for changing the light of the signal.

Prashant Jadhav et al. [2] designed a traffic control system which is mainly for priority-based traffic clearance. At the traffic lane, a web camera is placed. It is used to capture images of the road on which traffic needs to be controlled. A processing is done on each image to count the number of cars. A warning of heavy traffic is

intimated if the number of cars exceeds the threshold value. The processed data from MATLAB is used by the controller to send the command to the traffic pole.

Parichita Basak et al. [3] suggested a system to create an intelligent traffic control. It uses different types of image processing algorithms. Cameras are used to detect the vehicles along the traffic light controller. Sober Edge Detection, Pewitt Edge Detection, Roberts Zerocross Edge Detection, and Canny Edge Detection are some of the algorithms used to process the image.

Tejashri Gadekar et al. [4] proposed a system where each traffic controller (intersection) contains an RFID reader. An RFID is placed at each lane to track the vehicles to passing through it. A database is maintained at each intersection point to store the information regarding the vehicles that passes it. It also records the timestamp and traffic light. A RFID-enabled device is placed at every vehicle. Every vehicle has its unique vehicle identification number (VIN). The number has three components: priority, type, and the vehicle number.

Linganagouda R et al. [5] simulated an automatic intelligent traffic control system. The system has two parts. It explains the problems faced by ambulances, RFID concept, etc. In this technique, IR transmitter and receiver are used to make the vehicles' lane green, thus preventing traffic congestion.

R.U. Yawle et al. [6] proposed a Smart traffic control system which combines sensors with embedded technology. The dividers consist of IR sensors to detect the vehicles. Bluetooth and GPS are placed at every individual vehicle to identify the vehicles. Traffic density is measured by the IR sensors. The green light duration is based on the network congestion. GSM module is used to send message to the control vehicle if stolen vehicle is identified. GSM is also used to inform the traffic controller if the ambulance approaches the junction.

Bharath Kumar Perumallal et al. [7] developed traffic and vehicle monitoring system based on Internet-of-things (IoT). This system addresses the issues such as traffic congestion, vehicle spotting, and early warnings regarding jams. Implementation of the system is done using ATMEGA series microcontroller for monitoring purpose and NodeMCU for tracking purpose. The tracking module is placed inside the car. This module reads the latitude and longitude of the locations through which the car is traversed. GPS module reads the latitude and longitude. All sides of the lane consist of IR sensors. The traffic density can be identified by finding the number of connection breaks in the IR sensor pairs.

Vipul Vilas Sawake G. H. Raisoni C et al. [8] developed a fuzzy logic controller system which is used to control traffic signal at multiple intersections model. The fuzzy traffic light controller manages the congestion better than fixed time duration (static) traffic control system. The fuzzy traffic control system minimizes the vehicle waiting time at red light. By applying some fuzzy rules, it can extend the current green phase by adding different time duration.

Ashok. P.V et al. [9] developed an IoT-based traffic signaling system which consists of ultrasonic sensors placed at every 50 meters of the road. The ultrasonic sensors are used to capture the traffic density and communicate to Arduino for changing the traffic signal accordingly. The sensed information is sent to Pi3 using Wifi Module. With the help of data collected, analysis is made on heavy traffic and

less traffic with date and time and the same is communicated to web page of cloud. Further analysis is made on the data stored in cloud.

Vinidha Roc et al. [10] implemented a system which is used to control the traffic and helps for the emergency vehicles to reach at their destination easily and quickly. A system of cameras is used to capture the image of passing vehicles and to regulate traffic. Each camera captures the image and coordinates with other cameras to change the signal in a specific direction. Sound sensors are placed at the junction to detect the emergency vehicles. Sound sensors coordinate with the microcontroller and make the particular lane free.

K Senthilkumar et al. [11] proposed an approach to regulate traffic with the use of image processing and MATLAB simulation. This technique uses the observed images to compare with original images of the street taken in order to determine the traffic-level percentage and set the timing for the traffic signal. This technique solves the problems of real-life scenarios in the streets, thus enriching the traffic lights by adding image receivers like HD cameras and image processors. Then the input is imported into MATLAB and an algorithm is used for calculating the traffic on roads.

Swarup Kulkarni et al. [12] proposed a system for identifying the emergency vehicles like ambulance, police vehicle, and fire vehicle. In this system, RFID is used to identify the authenticated vehicle owner and stolen vehicle. The system has a buffer which will give a beep if the theft is detected. GSM technology is used to send message to the police and GPS is used to send the exact location of the theft.

Pampa Sadhukhan et al. [13] proposed an intelligent traffic control system which is based on Internet-of-things (IoT) that dynamically sets the signal operation time based on the measured values of traffic congestion density. The system consists of two modules such as traffic density monitoring module and traffic management module. The system measures the density of traffic congestion created at road crossings by using ultrasonic sensor node (USN) and then dynamically sets the signal operation time based on the measured values of traffic congestion density.

Eltayeb, A. S et al. [14] proposed a technology which uses GPS to indicate the location of the emergency vehicle. There are two modules: emergency vehicle module and traffic light module. This is an automatic preemption traffic control system, based on the global positioning system satellites. The system ensures the arrival of emergency vehicles to their destinations on the minimum time possible. Global positioning system is highly available and accurate. The system uses two metrics namely estimated Time to Arrive (ETA) and estimated critical time (CT).

Asaduzzaman, M et al. [15] suggested a system which uses priority signal control algorithm and a precedence graph. It uses eight-phase dual-ring controller. Two algorithms are implemented namely Single request algorithm and multiple emergency requests. The algorithms use the logic of greedy approach and depth first search.

Nellore, K et al. [16] proposed a technique which uses Priority for Emergency Traffic-based MAC (PE MAC) Protocol. It uses acoustic sensors, video capture from cameras at the intersection. The system uses three distance metrics: Euclidean

distance, Manhattan distance, and Canberra distance. Camera with image-processing techniques are used to process the captured camera images.

Meghana, B. S et al. [17] suggested a system which uses radio frequency identification (RFID) and analytics for real-time implementation. RFID tag stores details such as the vehicle number, vehicle type, and owner's details. The RFID antenna detects the tags using ultra high frequency (UHF) frequency band. This data is read by the RFID reader module and is sent to the server for further processing and analysis. GSM 4G connection accessed by Raspberry Pi 3 processor.

Rezaei et al. [18] proposed a system where a Hybrid FOG–Cloud intelligent control system called ReFOCUS is used. It is able to dynamically compute the best path for drivers those who are in or will be in the congested area, based on the current traffic density of different regions with considering the road future congestion status. There are two wireless communication technologies used. They are dedicated short-range communications (DSRC) cellular network technologies.

Amir et al. [19] proposed a RFID sensor-based emergency signal detection. Programmable logic controller (PLC) is used. It also uses higher priority encoder for priority detection.

El-Dalil et al. [20] proposed a priority level mutualism for emergency vehicle (PLMEV) algorithm. The system uses a GPS and radio frequency (RF) non-line of sight communication.

Al-Ostath et al. [21] proposed an emergency traffic light control system which uses the strobe light preemption system (preemption detector in line of sight), mobile infrared transmitter (MIRT) system (IR code more secure), acoustic systems, RF systems, GPS systems. The system provides a smooth flow for emergency vehicles such as ambulances to reach their destinations in time and thus minimizing the delay caused by traffic jams.

Shankar et al. [22] proposed a system which uses RFID Readers, Tags, and GPS. Ambulance detection uses the ambulance's position (D) and velocity (V) obtained from GPS module placed in the ambulance. When the ambulance is a few minutes away from the traffic signal an interrupt is raised in the distributed server aboard the upcoming traffic signal along with the calculation of the ambulance's arrival time.

The existing system stores the traffic congestion data in cloud which leads to higher latency for transmission. Some systems use cameras and web cameras to detect the vehicles on the path. Capturing of images and processing those images take more time. But the proposed system overcomes these drawbacks as it uses ZigBee transmitters in vehicles and ZigBee receivers in traffic light controller to detect the presence of vehicle on the lane. Our system uses the technique of end-device processing. So, it reduces the delay in transmission.

65.3 Proposed Work

The dynamic traffic light scheduler resolves the problem of traffic congestion of emergency vehicles. In this system ZigBee receiver and transmitters are used to send the priority and distance of the emergency vehicle to the traffic pole. The receiver at the traffic pole receives the signal and processes the requests based on higher priority to lower priority and from shorter distance to longer distance. This dynamic scheduling will reduce the traffic congestion. In this, processing is done at the end device. The data is not transferred to cloud so it reduces latency.

65.3.1 XBee Module S2C

The XBee S2C is an RF module and has transceiver, which gives a way to use RF communication at 2.4 GHz. Every S2C is equipped with the Arduino microcontroller which contains unique identification number (UIN). This UIN is based on the registration number of the vehicle. The communication between the transmitter and receiver is serial. It does not require additional hardware for transmission. The same module can be used as both transmitter and receiver because it is a transceiver. The microcontroller and S2C always communicate with serial communication. Rx pin of S2C is connected to Tx of microcontroller and Tx pin of S2C is connected to Rx pin of microcontroller. The data are transmitted and received at 9600 baud rate (Fig. 65.1).

Fig. 65.1 XBee module

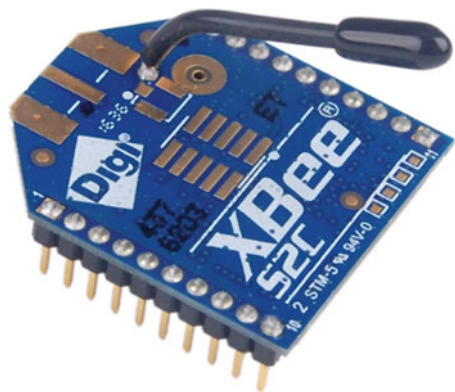


Fig. 65.2 Arduino Uno

65.3.2 *Arduino*

Arduino Uno is a microcontroller board based on 8-bit ATmega328P microcontroller. It also consists of other components such as crystal oscillator, voltage regulator, and serial communication along with ATmega328P. There are 14 digital input/output pins (out of which six can be used as pulse width modulation [PWM] outputs), six analog input pins, a USB connection. It also consists of a power barrel jack, an in-circuit serial programming (ICSP) header, and a reset button (Fig. 65.2).

65.4 Working Model

The system consists of various modules namely ZigBee module, SD card module, etc. The ZigBee transmitter is placed at the emergency vehicle and the receiver is placed at the traffic pole. The ZigBee transmitter sends the vehicle ID, priority and distance. The priority is assigned in such a way that ambulance is assigned a priority 1, fire service to priority 2, and police van to priority 3. Based on the priority, the request is processed. In the case of same priority, the distance from the transmitter to receiver is calculated, and based on the shortest distance the request is processed (Fig. 65.3).

The end devices involved in processing are called as fog nodes or fog devices. A list of registered emergency vehicles will be maintained at the end devices for validation. If an unregistered or new vehicle passes the lane, the request will be sent to the cloud for validation (Fig. 65.4).

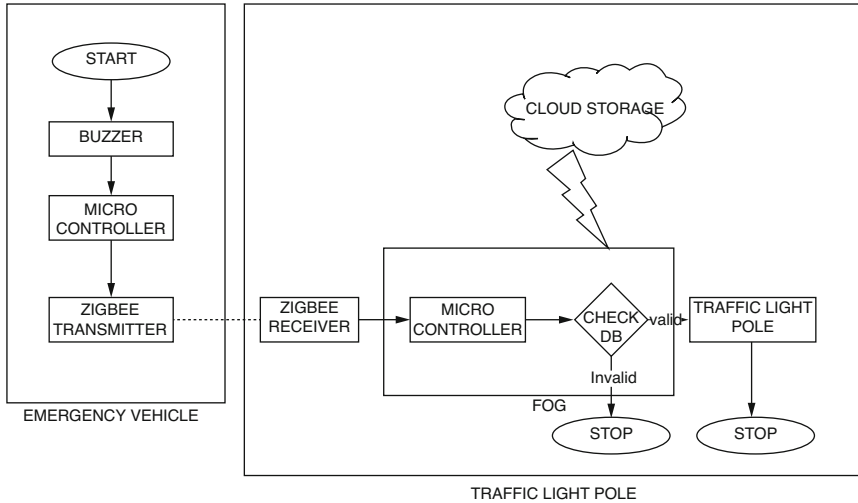


Fig. 65.3 System flow

Ambulance	Fire Service	Police	Normal Vechile
Priority : 1	Priority : 2	Priority : 3	Priority : 4
			

Fig. 65.4 Priority of emergency vehicles

65.4.1 Function at Emergency Vehicle

In this module, there is a ZigBee transmitter and a microcontroller in the emergency vehicle which gets activated when the buzzer is on. The ZigBee transmitter sends the data (ID, Priority, and Distance) to the traffic pole. The traffic pole consists of a ZigBee receiver and a microcontroller which receives the request from multiple transmitters placed inside emergency vehicles. The received requests are then processed (Fig. 65.5).

Fig. 65.5 SD module



Fig. 65.6 Transmitter 1

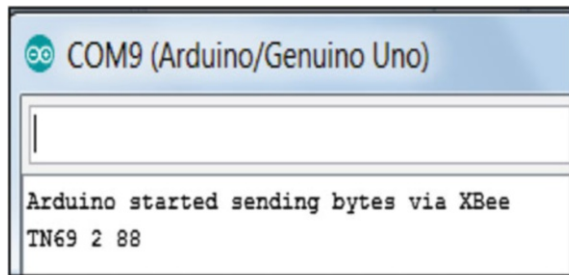
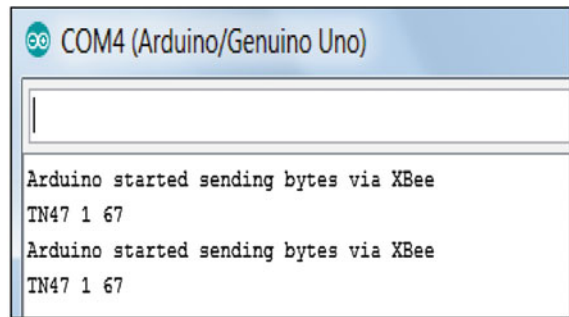
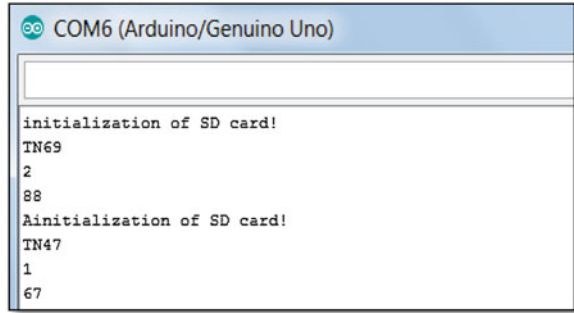


Fig. 65.7 Transmitter 2



65.4.2 Function at Traffic Pole

The ZigBee receiver receives the data from the transmitters, that is, emergency vehicles and sort the requests based on priority. If there is same priority then distance is considered for sorting. Distance calculation is done using round-trip time calculation. The timestamp at which the data is sent and received is noted and based on the time the distance between the transmitter and receiver is calculated (Figs. 65.6, 65.7 and 65.8).

Fig. 65.8 Receiver


```

COM6 (Arduino/Genuino Uno)

initialization of SD card!
TN69
2
88
Ainitialization of SD card!
TN47
1
67

```

There are different cases encountered for processing of request. They are as follows:

- Case 1: Only one emergency vehicle is present in a lane.
- Case 2: Two emergency vehicles, each having different priority, present in two different lanes.
- Case 3: Two emergency vehicles each having same priority (shortest distance).
- Case 4: Four emergency vehicles (all different priorities) placed at each of the four lanes (priority-based).
- Case 5: Four emergency vehicles (all have same priorities) placed at each of the four lanes (shortest distance).
- Case 6: Emergency vehicles are always arriving in particular two-phase only (periodic intervals for other two phase).

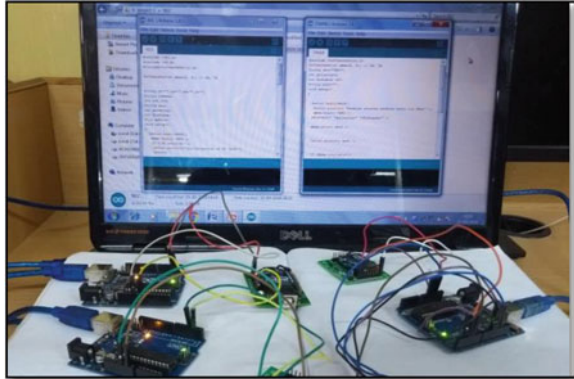
65.4.3 SD Card Module

Serial peripheral interface is used for communication between the microcontroller and the SD card. Communication takes place on digital pins 11, 12, and 13 (on most Arduino boards) or 50, 51, and 52 (Arduino Mega). SD card is selected by an additional pin. This additional pin can be the hardware SS pin – pin 10 (on most Arduino boards) or pin 53 (on the Mega) – or another pin specified in the call to `SD.begin()`. There are some basic inbuilt functions like `SD.begin()`, `SD.exists()`, `SD.mkdir()`, `SD.open()`, `SD.remove()`, `SD.rmdir()`.

65.5 Implementation Details

This section gives the details about the implementation details of the system. The emergency vehicle consists of a transmitter which sends the request (ID, Priority, and Distance) to the traffic light controller. The traffic light controller processes the request based on priority and distance (Fig. 65.9).

Fig. 65.9 System implementation



65.6 Conclusion

Fog computing complements the cloud to handle the large amount of data generated daily from the Internet of Things. The processing of data at the place where it is produced and needed solves the challenges of exploding data volume, variety, and velocity. Fog computing creates awareness and response to events by reducing the round trip to the cloud for analysis. With dynamic traffic light scheduling, loss of life due to traffic congestion on roads can be reduced. Faster processing of data leads to increased efficiency of the system. Our proposed system focuses on the vehicles around 100 meters, but this can be extended for larger distance by using protocols like LoRaWAN.

References

1. Sundar R, Hebbar S, Golla V (2015) Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. *IEEE Sens J* 15 (2):1109–1113
2. Jadhav P, Kelkar P, Patil K, Thorat S (2016) Smart traffic control system using image processing. *Int Res J Eng Technol (IRJET)* 3(3):1207–1211
3. Basak P, Kaur R (2016) Intelligent traffic control system using image processing. *Int J Sci Res (IJSR)* 5(8):1396–1398
4. Gadekar T, PriyankaChavare KC, Togrikar PS (2016) Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. *Imp J Interdiscip Res (IJIR)* 2(4):1041–1044
5. Linganagouda R, Raju P, Patil A (2016) Automatic intelligent traffic control system. *Int J Adv Res Electr Electron Instrum Eng* 5(7):5902–5906
6. Yawle RU, Modak KK, Shivshette PS, Vhaval SS (2016) Smart traffic control system. *SSRG Int J Electr Commun Eng* 3(3):22–25
7. Bharath Kumar Perumallal M, Babu S (2016) An intelligent traffic and vehicle monitoring system using internet of things architecture. *Int J Sci Res (IJSR)* 5(11):853–856
8. Sawake VV, Borkar P (2017) Traffic signal timings optimization using fuzzy logic controller. *Int J Adv Res Comput Sci Softw Eng* 7(4):172–177

9. Ashok PV, SivaSankari S, Mani V, Sankaranarayanan S (2017) IoT based traffic signaling system. *Int J Appl Eng Res* 12(19):8264–8269
10. Vinidha Roc A, Banuprakash PR, PaulAsir Nixon Raj G, Prasad L (2017) Smart traffic light systems. *Int J Emerg Res Manag Technol* 6(7):243–250
11. Senthilkumar K et al (2017) “Traffic analysis and control using image processing”, IOP conference series: materials science and engineering, vol 263, pp 1–5
12. SwarupKulkarni RA (2017) Intelligent traffic control system implementation for traffic violation control, congestion control and stolen vehicle detection. *Intelligent Traffic Control System Implementation for Traffic Violation Control, Congestion* 5(2):57–71
13. Sadhukhan P, Gazi F (2018, February) An IoT based intelligent traffic congestion control system for road crossings. 2018 International Conference on Communication, Computing & Internet of Things (IC3IoT 2018), S.R.M Engineering College, Chennai, India
14. Eltayeb AS, Almubarak HO, Attia TA (2013, August) A GPS based traffic light pre-emption control system for emergency vehicles. In: *Computing, Electrical and Electronics Engineering (ICCEEE), 2013 I International Conference on*. IEEE, pp 724–729
15. Asaduzzaman M, Vidyasankar K (2017, September) A priority algorithm to control the traffic signal for emergency vehicles. In: *Vehicular Technology Conference (VTC-Fall), 2017 IEEE 86th*. IEEE, pp 1–7
16. Nellore K, Hancke G (2016) Traffic management for emergency vehicle priority based on visual sensing. *Sensors* 16(11):1892
17. Meghana BS, Kumari S, Pushphavathi TP (2017, April) Comprehensive traffic management system: real-time traffic data analysis using RFID. In: *Electronics, Communication and Aerospace Technology (ICECA), 2017 international conference of*, vol 2. IEEE, pp 168–171
18. Rezaei M, Noori H, Rahbari D, Nickray M (2017, December) ReFOCUS: a hybrid fog-cloud based intelligent traffic re-routing system. In: *Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on*. IEEE, pp 0992–0998
19. Amir S, Kamal MS, Khan SS, Salam KA (2017, July) PLC based traffic control system with emergency vehicle detection and management. In: *Intelligent Computing Instrumentation and Control Technologies (ICICT), 2017 international conference on*. IEEE, pp 1467–1472
20. El-Dalil A, Sharkas M, Khedr M (2017, June) Priority level mutualism for emergency vehicle using game theory. In: *Vehicular Electronics and Safety (ICVES), 2017 IEEE International Conference on*. IEEE, pp 75–80
21. Al-Ostath N, Selityn F, Al-Roudhan Z, El-Abd M (2015, July) Implementation of an emergency vehicle to traffic lights communication system. In: *New Technologies, Mobility and Security (NTMS), 2015 7th International Conference on*. IEEE, pp 1–5
22. Shankar V, Gautham R (2015, March) Automated traffic signal for hassle free movement of ambulance. In: *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on*. IEEE, pp 1–5

Chapter 66

Cloud Database – A Technical Review



S. Sakthivel and J. Gnana Jayanthi

Abstract Recently, cloud computing technology has attracted several business organizations due to its wide delivery of computing resources as services. As a result, a service computing has emerged from the technology foundation perspectives such as Service-Oriented Architecture (SOA) and virtualizations of software and hardware. The increasing needs have database, information and expert knowledge systems have included vital role in cloud computing. Several issues and challenges are still to be focused and addressed with solution in cloud. Henceforth, this chapter is associated with developers, scientific persons, and users by highlighting and presenting the challenges and issues of cloud database.

Keywords Cloud computing · cloud database · cloud models · cloud architecture · cloud database challenges

Abbreviations

SOA	Service-oriented architecture
DBMS	Database management system
IaaS	Infrastructure as a service
PaaS	Platform as a service
SaaS	Software as a service
VM	Virtual machine
DaaS	Development as a service
SQLMR	Structure Query Language MapReduce
SOQL	Saleforce object query language
EC2	Amazon elastic compute cloud
DbaaS	DataBase as a service

S. Sakthivel (✉) · J. Gnana Jayanthi
PG and Research Department of Computer Science, Rajah Serfoji Government College,
Thanjavur, Tamil Nadu, India

CLI	Command line interface
API	Application programming interface
QoS	Quality of service

66.1 Introduction

Cloud computing is the one of most recent trends in distributed computing. With cloud computing, a new version of the Internet blooms with several significant features such as (i) accessing and sharing data and information with high speed and accuracy; (ii) prevention of software piracy forever; (iii) reduction in cost of conducting business; (iv) monitoring the activity of employees on the application in real time; (v) automation in product development, distribution, and management etc. Rackspace, Amazon, Oracle, Google are giant IT companies offering cloud services [1]. Henceforth, adopting cloud computing creates new opportunities and makes people life easier and fast.

However, majority of cloud applications are data driven, Database Management System (DBMS) software powering several applications from their critical phases under the cloud environment. But, traditional, DBMS software's are not compatible with cloud friendly, because cloud services such as web servers and application servers scale from few machines to hundreds. Existing databases are resulting in a heavy performance impact during elastic scaling. In cloud computing environment, DBMS must follow the properties such as scalability, elasticity, autonomic, and load balancing in addition to the other properties like high-level functionality, consistency, performance, and reliability [2].

This chapter reviews cloud computing and cloud database with some of the desired properties such as scalability, elasticity, autonomy, reliability, availability, and consistency; and is organized as follows. Section 66.2 overviews the cloud models. A summary on the technologies that enable cloud computing is presented in Sect. 66.3. Section 66.4 briefs the cloud architecture. Cloud security is presented in Sect. 66.5. A brief overview about cloud database is presented in Sect. 66.6. Section 66.7 outlines the methodology of database security. Section 66.8 briefly describes the challenges and issues of cloud database.

66.2 Cloud Models

Recently, several organizations have identified the benefits of using cloud environment. Understanding the needs of the organization, cloud models are categorized into two as (i) deployment model, and (ii) service model [1, 2].

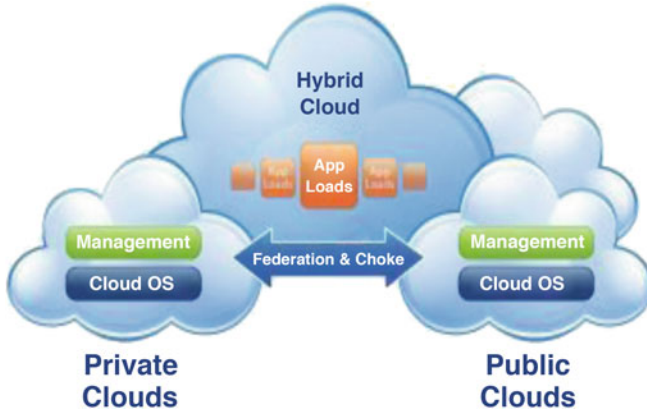


Fig. 66.1 Deployment models of the cloud

66.2.1 Deployment Model

The deployment model is further classified into three major clouds as (i) public, (ii) private and (iii) hybrid cloud.

Public Cloud: The cloud infrastructure is made available to the public or a large industry group and is owned by an organization selling cloud services.

Private Cloud: The cloud infrastructure dedicated to a single organization. It may be managed by the organization or a third party and may exist on or off premise.

Hybrid Cloud: The cloud infrastructure is a composition of two or more cloud (public, private) that unique entities but are bound together by standardized or proprietary technology that enables data and application portability (Fig. 66.1).

66.2.2 Service Model

The cloud services are provided to the cloud users through three major certain service models such as (i) Infrastructure as a Service (IaaS), (ii) Platform as a Service (PaaS), and (iii) Software as a Service (SaaS).

Infrastructure as a Service (IaaS): Shared infrastructure such as servers, storage, and network are delivered as a service over the Internet. Some examples include Amazon Web Services, Rackspace Cloud.

Platform as a Service (PaaS): Application development framework offered as a service to developers for quick deployment of their code. Some examples for Platform as a Service (PaaS) include Google App Engine, Hurok, and Cloud Foundry.

Software as a Service (SaaS): Application software are offered as a service using a multitenant model which can be consumed using web browsers. Gmail, Salesforce, etc. are some of the examples for implementing SaaS.

66.3 Enabling Technologies

Several other computing technologies are supporting for the booming of cloud computing and some of the noteworthy technologies are briefed out here [3].

Grid Computing: Grid computing, a distributed computing paradigm, coordinates networked resources to achieve common computational objectives. The development of Grid computing was originally driven by scientific applications which are usually computation-intensive. Cloud computing is similar to Grid computing in that it also employs distributed resources to achieve application-level objectives. However, cloud computing takes one-step further by leveraging virtualization technologies at multiple levels (hardware and application platform) to realize resource sharing and dynamic resource provisioning [4] (Fig. 66.2).

Utility Computing: Utility computing represents the model of providing resources on-demand and charging customers based on usage rather than a flat rate. Cloud computing can be perceived as a realization of utility computing. It adopts a utility-based pricing scheme entirely for economic reasons. With on-demand resource provisioning and utility-based pricing, service providers can truly maximize resource utilization and minimize their operating costs [3, 4].

Virtualization: Virtualization is a technology that abstracts away the details of physical hardware and provides virtualized resources for high-level applications. A virtualized server is commonly called a virtual machine (VM). Virtualization forms the foundation of cloud computing, as it provides the capability of pooling

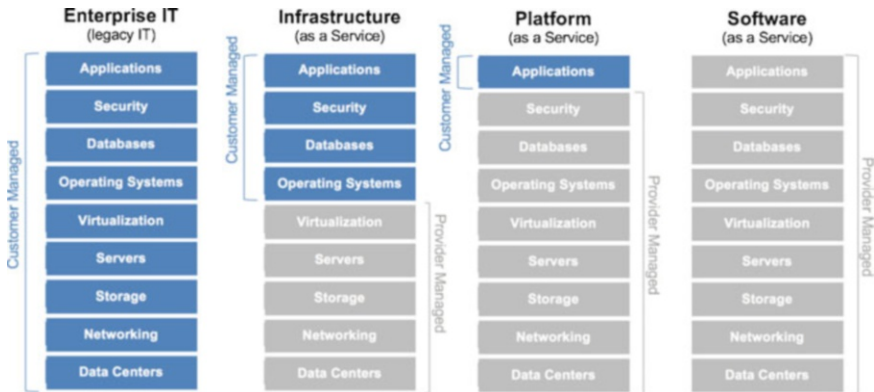


Fig. 66.2 Service models of the cloud

computing resources from clusters of servers and dynamically assigning or reassigning virtual resources to applications on-demand [5].

Autonomic Computing: Originally coined by IBM in 2001, autonomic computing aims at building computing systems capable of self-management, that is, reacting to internal and external observations without human intervention. The goal of autonomic computing is to overcome the management complexity of today's computer systems. Although cloud computing exhibits certain autonomic features such as automatic resource provisioning, its objective is to lower the resource cost rather than to reduce system complexity [6].

Therefore, cloud computing leverages virtualization technology to achieve the goal of providing computing resources as a utility. It shares certain aspects with grid computing and autonomic computing, but differs from them in other aspects. Therefore, it offers unique benefits and imposes distinctive challenges to meet its requirements.

66.4 Architecture of Cloud

Cloud Architecture refers to the various components in terms of databases, software capabilities, applications, etc., engineered to leverage the power of cloud resources to solve business problems. Cloud architecture defines the components as well as the relationships between them. The various components of cloud architecture are: (i) on premise resources, (ii) cloud resources, (iii) software components and services, (iv) middleware [7].

The entire cloud architecture is aimed at providing the users with high bandwidth, allowing users to have uninterrupted access to data and applications, on-demand agile network with possibility to move quickly and efficiently between servers or even between clouds and most importantly network security.

The various cloud-based services have their own distinct and unique cloud architectures: (i) Software as a Service (SaaS) involves software hosted and maintained on Internet. With SaaS, users do not have to install the software locally. (ii) Development as a Service (DaaS) involves web-based development tools shared across communities. (iii) Platform as a Service (PaaS) provides users with application platforms and databases, equivalent to middleware services. (iv) Infrastructure as a Service (IaaS) provides for infrastructure and hardware such as servers, networks, and storage devices running in the cloud, available to users against a pay per usage basis. A general cloud architecture comprising (i) external layer, (ii) conceptual middleware layer, (iii) conceptual layer, (iv) physical middleware layer, and (v) physical layer.

External Layer: This layer is presentable to the users. More and more companies are moving their data management applications from expensive, high-end servers to the cloud that is composed of cheaper, commodity machines.

Conceptual Middleware Layer: This layer provides the facility of hiding the conceptual level heterogeneity among different databases like SQL, DB2, and Oracle.

Conceptual Layer: As cloud deals with various types of data, here users need to combine the traditional data with the data that are placed on the cloud, so various types of systems are required for Cloud Databases that provides all these functionalities. Some of the systems or languages like Big integrator, SQLMR, Salesforce Object Query Language (SOQL).

Physical Middleware Layer: This layer provides the facility of hiding the heterogeneity across the different platforms like windows, Mac OS, Linux.

Physical Layer: This layer represents the physical representation of the data. In a cloud database service, the backend is being overseen by a physical layer that is responsible for the continuous monitoring and configuring of the database to achieve optimal scaling, high availability, multitenancy, and effective resource allocation in the cloud.

66.5 Cloud Database

Data management applications are potential candidates for deployment in the cloud. IT Companies such as Google, Microsoft, [Salesforce.com](https://www.salesforce.com), Rackspace, Amazon Elastic Compute Cloud (EC2) etc. are offering cloud database in the computing environment [8, 9].

DataBase as a Service (DBAAS) is the DBMS outsourcing. It is the most required cloud services. The data transmission and the data management are the most considerable case in the sense cloud database. Data management more costly compared with data transfer. To reduce the cost and maintenance of the data management, the best way is outsourcing the DBMS through cloud. Cloud databases offer a freedom to choose SQL variants. Cloud databases include support for multiple My Structured Query Language (MYSQL) compatible variants and versions, including MYSQL.

Cloud databases is customized to deliver fast, predictable performance for high I/O MYSQL database workloads. Cloud databases instances come preconfigured for peak performance. Customers can also customize the default configuration settings and make persistent changes to further optimize for specific workload using the Command Line Interface (CLI) and Application Programming Interface (API). Part of what allows Cloud databases to create this high-performance infrastructure is container-based virtualization. Each Cloud databases instance also comes with an attached storage volume.

66.6 Rationale and Significance of the Cloud Databases Study

66.6.1 Cloud DBMS Wish List

Cloud computing has become the norm of the today's computing world. This paradigm shift is looking for the databases that are compatible in all respect. Hence, in 2009, Daniel Abadi introduced the Cloud DBMS wish list that looks for the following characteristics [10]:

1. *Efficiency*: The pricing model based on the consumption of the required storage, used network bandwidth and the computational power.
2. *Fault Tolerance*: Transactional as well as analytical DBMS that supports the complex query evaluation without compromising the ACID properties of DBMS.
3. *Ability to Run in Heterogeneous Environment*: Cloud computing is essentially a system working in the heterogeneous environment and inconsistent or the degraded performance of this hampers the query execution.
4. *Ability to Operate on Encrypted Data*: As customers of cloud are relying for storage, manipulation, and retrieval of sensitive data through the different service models available, the data encryption and decryption capability is the mandate.
5. *Ability to Interface with Business Intelligence*: Different customized tools will be used for data visualization and the appropriate and efficient interfaces are needed to satisfy the queries.
6. *Security*: Data especially in the public cloud must be secured enough without compromising on to the national legislation policies for data storage.

66.6.2 Cloud Service Provider Challenges

Recent advancements in cloud computing are typically posing two major challenges for the cloud service providers:

- data security
- query optimization

The cloud users are vulnerable to different data security breaches as these users are establishing their identity every time they use the cloud service. While doing so, they end up providing sensitive personal information (e.g., name, address, credit card details, phone number, organization details, etc.). Every cloud service utilization by the customer is leaving the footprints with the trail of personal information available at a click event. If not being protected properly, may be misused.

The use of technology in different spheres are increasing the challenges for security and expecting the scalable support for the queries reaching the cloud. These shared resources are being used to process the diverse requests from the

cloud customers who potentially have widely different expectations from the services being offered by the cloud environment.

66.7 Database Infrastructures as a Service

66.7.1 DbaaS

DbaaS is attractive for two reasons [11]

1. Due to economies of scale, the hardware and energy costs incurred by the users are likely to be much lower when they are paying for a share of a service rather than running everything themselves.
2. The costs incurred in a well-designed DBaaS will be proportional to the actual usage (“pay-per-use”)—this applies to both software licensing and administrative costs.
3. By centralizing and automating many database management tasks, a DBaaS can substantially reduce operational costs and perform well.

66.7.2 Query Optimization in Cloud Databases

The third-party search services which are providing the intermediate results are composed of the arbitrary operations on the data that may have unknown semantics. The algebraic properties of the data stored and manipulated vary with the performance characteristics of the cloud environment [12]. As in relational algebra that has been the basis of traditional DBMS, there is no fixed set of operations that need to be performed on the cloud data to get the desired query result. The cloud service providers QoS constraints to add the following: resource utilization cost, data availability, and consistency of the database etc.

The data processing graphs which are the results of the cloud query processing are flexible and not fixed a priori. The queries coming from the cloud consumers approach with varying constraints and scalability, making data flow optimization a challenging problem in the cloud databases. This also requires dynamism to be embedded in query execution.

66.8 Cloud Database – Challenges and Issues

Scalability: The lack of cloud computing features to support RDBMSs associated with enterprise solutions has made RDBMSs less attractive for the deployment of large-scale applications in the cloud. This drawback has resulted in the popularity of NoSQL [13].

Availability: Availability refers to the resources of the system accessible on-demand by an authorized individual [14]. In addition, with the increasing number of cloud users, cloud service providers must address the issue of making the requested data available to users to deliver high-quality services. [15]

Autonomic: The cloud database management must have a property like autonomic or self-managing, which belongs to the scalability and elasticity.

Consistency: Data availability and durability is the main requirement of the cloud vendors. It can be achieved by maintaining certain number of replicas of the data at different locations. However, to maintain the consistency between these replicas become a complicated issue.

Duplication: Data duplication is one of the problems presented by distributed data storage on the cloud side. In general, for the purpose of data recovery, data are repeatedly stored on different servers that are geographically distributed.

Data integrity: One of the main challenges that must be addressed is to ensure the correctness of user data in the cloud. Given that users may not be physically able to access the data, the cloud should provide a mechanism for the user to check whether the data is maintained [16].

Data quality: Obtaining high-quality data from vast collections of data sources is a challenge. High-quality data in the cloud are characterized by the data consistency. If data from new sources are consistent with the data from other sources, then the new data are of high quality [17].

Heterogeneity: In a cloud environment, users can store data in structured, semi-structured, or unstructured format. Structured data formats are appropriate for today's database systems, whereas semi-structured data formats are appropriate only to some extent. Unstructured data are inappropriate [18], because they have a complex format that is difficult to represent in rows and columns. According to Kocarev and Jakimoski [19], the challenge is how to handle multiple data sources and types.

66.8.1 Open Research Issues

Data Staging: The most important open research issue regarding data staging is related to the heterogeneous nature of data. Data gathered from different sources do not have a structured format. For instance, mobile cloud-based applications, blogs, and social networking are inadequately structured similar to pieces of text messages, videos, and images. Transforming and cleaning such unstructured data before loading them into the warehouse for analysis are challenging tasks.

Distributed Storage Systems: Numerous solutions have been proposed to store and retrieve massive amounts of data. Some of these solutions have been applied in a cloud computing environment. However, several issues hinder the successful implementation of such solutions, including the capability of current cloud technologies to provide necessary capacity and high performance to address massive amounts of data [20], optimization of existing file systems for the

volumes demanded by data-mining applications, and how data can be stored in such a manner that they can be easily retrieved and migrated between servers.

Data Analysis: The speed of stream data arriving from different data sources must be processed and compared with historical information within a certain period of time. Such data sources may contain different formats, which makes the integration of multiple sources for analysis a complex task [21].

66.9 Conclusion

The increased use of the database infrastructure as a service in this cloud computing era has put forth many challenges pertaining to the organization and management of the data storage. This leads to many constraints being applied onto cloud database query processing, security, heterogeneities, data staging, data analysis etc., some of the new techniques have to implement to improve the cloud database performance.

References

1. Buyya R, Yeo CS, Venugopal S, Broberg J, Brandic I (2009) Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. *Future Gener Comput Syst* 25(6):599–616. Elsevier Science, Amsterdam
2. Agrawal D, Abbadi AE, Das S, Elmore AJ (2011) Database scalability, elasticity and autonomy in cloud. In: *The Springer proceeding of the international conference on database systems for advanced applications*, vol 6587. Springer Veering LNCSB, pp 2–15
3. Dias de Assunção M, Buyya R, Venugopal S (2008) InterGrid: a case for internetworking islands of grids. *Concurr Comput Pract Exp* 20(8):997–1024. ISSN:1532-0626, Wiley Press, New York
4. Erwin DW, Snelling DF (2004) UNICORE: a grid computing environment. In: *Proceedings of the 7th international Euro-Par conference on parallel processing (Euro-Par '01)*, Manchester, UK, vol 3032. Springer, Berlin, pp 495–502
5. Takabi H, Joshi JBD, Ahn G (2010) Security and privacy challenges in cloud computing environments. *IEEE Secur Priv* 8(6):24–31
6. Agrawal D, Abbadi AE (2011) Albatross: lightweight elasticity in share storage databases for the cloud using live data migration. In: *Proceedings of VLDB endowment*, vol 4, p 8
7. Ko Rayan KL (2011) Trust cloud: a framework for accountability and trust in cloud computing. Additional publication information: to be published and presented at the 2nd IEEE cloud Forum for Practitioners (IEEE ICFP 2011), Washington DC, USA, July 7–8
8. Santos N, Gummadi KP, Rodrigues R (2009) Towards trusted cloud computing. In: *The proceedings of the conference on Hot topics in cloud computing (HotCloud'09)*. USENIX Association, Berkeley
9. Broberg J, Buyya R, Tari (2009) MetaCDN: harnessing 'Storage clouds' for high performance content delivery. *J Netw Comput Appl* 32(5):1012–1022. Elsevier, Amsterdam
10. Abadi DJ (2009) Data management in the cloud: limitations and opportunities. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*
11. Curino C et al (2011) Relational cloud: a database-as-a-service for the cloud. *Dspace@MIT*
12. Nurmi D, Wolski R et al. The Eucalyptus open-source cloud computing system

13. Cattell R (2011) Scalable SQL and NoSQL datastores, *ACMSIGMOD record*, 39(4), ACM New York, 2011, 12–27
14. Zissis D, Lekkas D (2012) Addressing cloud computing security issues. *Future Gener Comput Syst* 28:583–592
15. Lee S, Park H, Shin Y (2012) Cloud computing availability: multi-clouds for bigdata service. *Commun Comput Inf Sci* 310:799–806
16. Sravan Kumar R, Saxena A (2011) Data integrity proofs in cloud storage. In: *Proceedings of the third international conference on communication systems and networks (COMSNETS)*, pp 1–4
17. Weber K, Rincon G, Van Eenennaam A, Golden B, Medrano J (2012) Differences in allele frequency distribution of bovine high-density genotyping platforms in holsteins and jerseys. *Western Section American Society of Animal science*, p 70
18. Che D, Safran M, Peng Z (2013) From bigdata to bigdata mining: challenges, issues, and opportunities. In: Hong B, Meng X, Chen L, Winiwarter W, Song W (eds) *Database systems for advanced applications*. Springer, Berlin/Heidelberg, pp 1–15
19. Kocarev L, Jakimoski G (2001) Logistic map as a block encryption algorithm. *Phys Lett* 289 (4–5):199–206
20. Leavitt N (2013) Storage challenge: where will all that bigdata go? *Computer* 46:22–25
21. Assuncao MD, Calheiros RN, Bianchi S, Netto MA, Buyya R (2013) Big data computing and clouds: challenges, solutions, and future directions, *arXiv preprint arXiv:1312.4722*

Chapter 67

Projection of Population and Prediction of Food Demand Through Mining and Forecasting Techniques



J. Antonita Shilpa and V. Bhanumathi

Abstract Food, clothing and shelter are the basic needs of man. The most essential among the needs is food. There is always a wide gap between supply and demand because of the changes in food preference. The change in food preference is the major factor in prediction of food demand. The proposed method uses the second-order Taylor series for the projection of population; having the estimated population as input as the food demand based on the change in food requisite is anticipated. The implementation is carried out through Java. The population and food demand of the continental U.S. are projected by the proposed method. The food demand prediction through the proposed method is similar to the actual demand with deviation close to 0.1%.

Keywords Curve fitting · Gaussian function · Intelligence continuum · Linear approximations · Maximum likelihood estimate · Relative growth · Taylor series

Abbreviations

A	Actual value
MLE	Maximum likelihood estimate
P	Predicted value
POP	Population
RGC	Relative growth coefficient
US	United States

J. Antonita Shilpa (✉) · V. Bhanumathi
Anna University Regional Campus, Coimbatore, Tamil Nadu, India

67.1 Introduction

Food is one of the most essential components for any living organism. The growth in lifestyle has led to various changes in behaviour, conduct, mannerism, dressing, education, communication formats, culture and most importantly food. The traditional kind of food framed by our ancestors had the reason of necessity and availability. The globalization concept has opened the doors for a variety of foods and also various food habits. The traditional method of food intake has been gradually changing for a few decades now. Though the change is not important, the adaptation of change is necessary, thus came the analysis of change in food habits.

As per an article on world agricultural supply and demand estimates by the United States Department of Agriculture [1], the food demand has always been higher than the supply, leading to increase in import. The food items that were more than the actual need were either exported or simply wasted. Natural ways of preserving food for future could be applied for a variety of items, but there were items that could not be preserved without a preservative, so it led to the increase in chemical intake. The food which was the basic need was produced, wherein the food with changing consumption was generally imported to avoid food wastage and intake of preservatives, but this paved way for the increase in the expense of food.

As per another article on climate impacts on agriculture and food supply by the United States Environmental Protection Agency [2], the nutrient contents in food consumption of the past century had been studied in detail based on the age, sex, lactation and pregnancy.

As per that article, nutrient content of the United States Food Supply, 1909–2000, by the United States Department of Agriculture [3], it is shown that the food intake had been gradually changing which would be a continuum even for the future, with no knowledge as of how the change would be. The proposed method projects food consumption up to 2100 accounting the data of change in food requisite from data system by Wenhe [4].

67.2 Related Work

67.2.1 *Taylor Series*

The linear approximation technique is the most common method with less deviation from the actual data. Adalgiza del Pilar Rios et al. [5] stated Polynomial Approximations for the processing, control and detection, but not projections. It was also discussed by Stephen B. Duffull et al. [6] that it was the best for stable intervals, for example, power consumption of a device per hour. The problem is that it cannot be used for data with unstable intervals. Youn-Kyou Lee et al. [7] explained

trigonometric function tangent with the first, second-order equations, which resulted in good performance.

67.2.2 Maximum Likelihood Estimation (MLE)

MLE is a method for estimating the relative growth coefficient of the data distribution for the anticipation of future data. The research concluded that it is an essential task in prediction in several ways. Divya Anand et al. [8] resolved ideas about Linear Discriminant Method, which used mean and covariance for a forecast of continuously changing data and found that it was more accurate than the Taylor series estimation; it was misleading when continuous variables are independent. Lavanya A. et al. [9] unravelled the theory of Inverse Likelihood which spotted the dissimilarities of data and was the best sharpening technique. From the study, we can conclude that estimating MLE is the basic step for projecting, but it cannot be used to envision over a long span as a cyclic process.

67.2.3 Gaussian Distribution

Gaussian function is a bell-shaped curve, used for estimating constant valued functions. The importance of the Gaussian function in various fields can be known through continued survey. A. F. Coşkun et al. [10] elucidated the overview of Integral Distribution, which is the curve fitting of several probability densities via MLE to attain maximum goodness of fit. Lazaros Iliadis et al. [11] put together the concepts of Multivariate Distribution for future forecasting using the Gaussian curve; that was the utmost preferred for future inference. Bassant Selim et al. [12] unravelled the mixture of Gaussian Distributions of maximization and expectation algorithms, which was simple and accurate and whose analytical format was similar to the generalized Gaussian distribution.

67.3 Proposed Method

The proposed method projects food consumption up to 2100 accounting the data of change in food requisite for the years from 1970 to 2014. The data for the population of the United States [13] from 1909 to 2014 are accounted for the population forecast. Though a century's data are the input of population, the prediction can be done only up to 2100, a span of 86 years (Fig. 67.1).

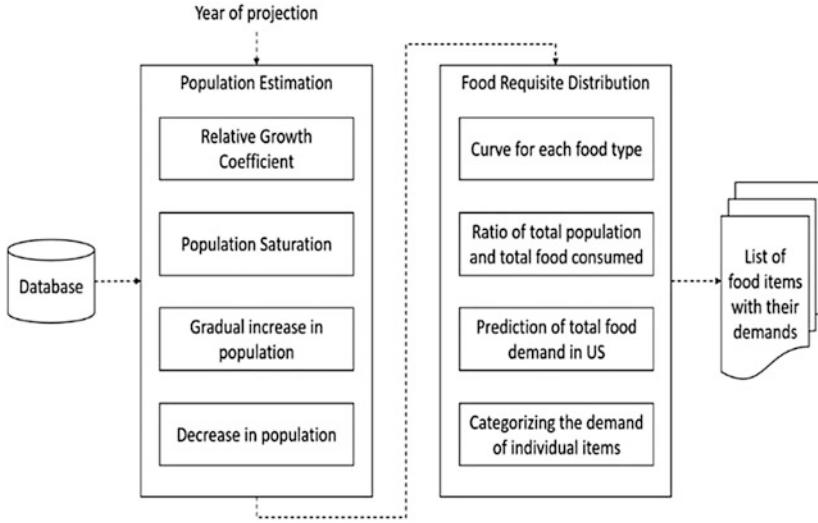


Fig. 67.1 Architectural diagram for the proposed system

67.3.1 Population Estimation

The population estimation is subdivided into relative growth coefficient, population saturation, gradual increase in population and decrease in population [14]. Relative growth coefficient (RGC) is estimated through the inverse MLE of the population data using the Taylor’s approximations. On solving the logistic curve equation, the population saturation is,

$$P_s = \frac{2P_0P_1P_2 - P_1^2(P_0 + P_2)}{P_0P_2 - P_1^2} \tag{67.1}$$

The equation for population projection is as follows,

$$POP_{proj} = POP_{last} + B (proj - last) \tag{67.2}$$

where

$$B = \frac{POP_{last} - POP_{first}}{last - first} \tag{67.3}$$

- POP_{first} = the population of the smallest year entry in the database
- POP_{last} = the population of the largest year entry in the database
- POP_{proj} = the population of the projected year
- $first$ = the smallest year entry in the database

last = the largest year entry in the database

proj = the projected year

The equation for population projection decrease after the saturation is,

$$POP_{proj} = RGC * (Ps - POP_{last}) * t \quad (67.4)$$

67.3.2 Food Requisite Distributions

The consumption of a specific food item does not remain constant over the years. The preference in food intake is a continuously changing process. By analysing the data, the change in intake of food is an important impact for production is clear. The total food ratio is the sum of all individual ratios. The individual predictions. The equation for $R_{p/ip}$ is,

$$R_{p/ip} = \frac{\sum \{pDemand_j | j = individual\ food\ item\}}{pDemand_{total}} \quad (67.5)$$

where

$pDemand_j$ = the demand predicted for food item j

$pDemand_{total}$ = the demand predicted for total food

The demand of individual food items, d_i is derived from the equation,

$$d_i = \frac{pDemand_i}{R_{p/ip}} \quad (67.6)$$

From Eq. 67.6, the demand of individual food items can be known with more accuracy. The equations are executed via programming modules written in Java.

67.4 Results and Discussion

The world's population has been continuously increasing for a very long time now; the population is expected to reach saturation in a few years. The anticipation can only be done for a few decades after saturation.

The proposed system has a strong mathematical foundation and the effects are seen through the accuracy in results in Table 67.1. The data consists of the US population; hence the future population of the US is only predicted. It is found that saturation will be around 2080 in the US. The deviation of population (POP) is only around 0.1%. The deviation of prediction is only around 0.1%. Table 67.1 contains

Table 67.1 Correctness of prediction to actual data

2015			2016			2017		
	P	A		P	A		P	A
POP	321.36	321.1	POP	321.36	321.1	POP	325.78	325.1
Apple	2847	2840	Apple	2847	2840	Apple	2886	2890
Banana	2579	2572	Banana	2579	2572	Banana	2614	2603
Beef	5758	5740	Beef	5758	5740	Beef	5836	5824
Chicken	8489	8478	Chicken	8489	8478	Chicken	8604	8589
Potato	1266	1267	Potato	1266	1267	Potato	1284	1268
Tomato	1877	1890	Tomato	1877	1890	Tomato	1903	1905

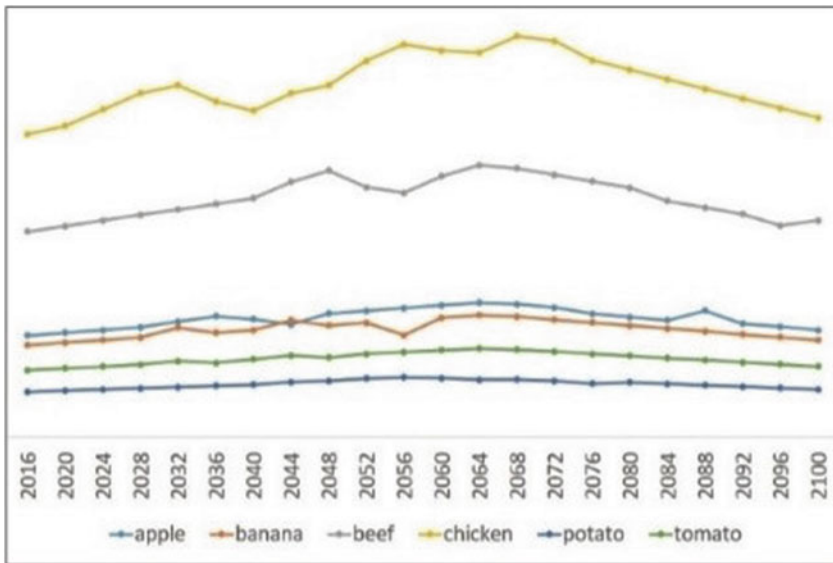


Fig. 67.2 Change in food preference representation

the data of approximation and actual data for 2015, 2016 and 2017, through which the correctness of approximations could be checked, since the data for these years are now available. On proving the correctness for 2015, 2016 and 2017, the stability of the system is found to be high and the deviation to actual is very less proven.

The analyses of the population are optimal. Figure 67.2 depicts the ratio between changes in food preferences. The prediction of Fig. 67.2 is derived from the input of the table 67.1. Chicken remains to be the highest food consumed. Even though there are peaks and valleys, it will maintain to be on top of the list. Beef consumption is slightly linear with less deviation.

Apple has a few peaks and valleys up to 2050, then it will tend to be linear until saturation. There is a change at the saturation in apple consumption. Banana possesses a wobble-natured curve from 2030 up to 2060. Then it is likely to be

linear. On granular levels, we can spot that there is fluctuation in the consumption of tomato and potato too. The minimal fluctuation in the curve is due to the very less intake compared to the intake of chicken. On having individual graphs for each food item, the clear flow of peaks and valleys is known.

The change in deviation is the gap between one curve and the other. For this dataset the deviation can be easily identified between apple and banana, since there is an overlap in the 2030s and 2040s and there is slight convergence after the saturation. There could be more such detections on considering more food items such as milk, grains, fat, cereals and others.

67.5 Conclusion

On seeing the outcome of the Java implementation, we can state that the proposed system is the best for the prediction of food accounting the change in preference. The approximation is made up only to 2100, but this cannot be determined as a continuum. When the continuum is followed in cycles, the slight deviation may turn out to cause massive deviations in the years to come. Hence the continuum is not possible and the years are limited to 2100. The change in approximation methods can be made, but the deviation of the proposed system is very less, hence the suggested approximation is as good as any of the methods which are not discussed. The extension of the proposed work may be done in various area, such as increasing the food items considered, prediction for larger population nations, prediction of decline in health due to changes in diet, etc.

References

1. Article on World Agricultural Supply and Demand Estimates by the United States Department of Agriculture, April 2015, May 2016, November 2017
2. Article on Climate Impacts on Agriculture and Food Supply, by United States Environmental Protection Agency, January 2017
3. Article on Nutrient Content of the U.S. Food Supply, 1909–2000 by the United States Department of Agriculture, Home Economics Research Report, November 2004
4. Source: The ERS Food Availability (Per Capita) Data System (FADS) Link: <https://data.world/wendyhe/the-changing-american-diet>
5. Adalgiza del Pilar Rios et al (2017) Taylor series approximation of ZIP model for on-line estimation of Residential Loads' parameters
6. Duffull SB et al (2017) Assessing robustness of designs for random effects parameters for nonlinear mixed-effects models. *J Pharmacokinet Pharmacodyn*
7. Youn-Kyou Leea et al (2017) Analytical representation of Mohr failure envelope approximating the generalized Hoek-Brown failure criterion. *Int J Rock Mech Min Sci*
8. Anand D et al (2017) Building an intelligent integrated method of gene selection for facioscapulohumeral muscular dystrophy diagnosis. *Int J Biomed Eng Technol* 24(3):285–296
9. Lavanya A et al (2016) Inverse maximum likelihood-based edge detection for segmentation of breast lesion using active contour. *Int J Biomed Eng Technol* 22(3):272–283

10. Coşkun AF et al (2017) A statistical-based examination on wind turbines' bi-static scatterings at 1 GHz frequency. *J Electromag Waves Appl* 32(3):347–362. <https://doi.org/10.1080/09205071.2017.1391127>
11. Iliadis SS et al (2010) A fuzzy inference system using Gaussian distribution curves for forest fire risk estimation. In: *IFIP international conference on artificial intelligence applications and innovations*
12. Selim B et al (2016) Modeling and analysis of wireless channels via the mixture of Gaussian distribution. *IEEE Trans Veh Technol* 65(10):8309–8321
13. *World Population and Human Capital in the Twenty-First Century* by Wolfgang Lutz et al. Oxford University Press, 2014
14. *A Primer of Real Analytic Functions* by Steven G. Krantz, Harold R. Parks, Second Edition Springer Science + Business Media, LLC

Chapter 68

Detection of Hairline Fracture Foot Using Canny Operator and Wavelet Packet Transform



D. S. Karthika, K. S. Biju, G. H. Silpa, and C. Girish Kumar

Abstract Bone fracture detection is the most common and usual technique in the medical imaging field. This detection is nowadays easy, since there exist various methodologies for diagnoses. However, detecting open fractures or severe fractures are easy by these detection techniques. While taking the hairline fractures, the case becomes complicated. It is not always possible to detect hairline fractures through bare eye. This is also a challenge for medical imaging. In this study, detection of hairline bone fracture is a problem to be solved via implementing Canny Operator and Wavelet Packet Transform. The preprocessed digital X-ray input image is transformed, and the fracture is detected using Canny Operator. Using Canny edge detection and Wavelet Packet Transform, hairline fractures can be detected in the earlier stage itself. The results are encouraging the effectiveness of Canny edge detection and four-level Wavelet Packet Decomposition is an image enhancer, which is proved in the study.

Keywords Canny edge detection · Hairline fracture · Wavelet packet transform · X-ray image

Abbreviations

CT Computed Tomography
WPT Wavelet Packet Transform

D. S. Karthika · G. H. Silpa
Electronics and Communication Department, Government Engineering College,
Thiruvananthapuram, India

K. S. Biju (✉)
Electronics and Communication Department, Government College of Engineering, Kannur,
India

C. Girish Kumar
Faculty of Engineering & Computer Technology, AIMST University, Kedah Darul Aman,
Malaysia

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_68

68.1 Introduction

The major part in the human skeleton system is bone. Similar to all other body parts, bone also undergoes damages. Damages of bone due to various cracks are known as bone fractures. These fractures can be subtle or extreme depending on the intensity of cracks. A fracture in the bone will crack a bone into various parts [1]. Bone can be fractured in crosswise, lengthwise or in multiple pieces. Common types of bone fractures are include stable fractures, open or compound fractures, transverse fractures, oblique fractures, comminuted fractures, and avulsion fractures. Certain fractures in foot include plastic bowing, buckle-type fracture, hairline fracture, etc. [2, 3].

Various medical imaging techniques are used for the diagnosis of bone fractures. X-ray technology is the common and most ancient technology for the imaging of bone fractures. The main principle behind it is electromagnetic radiation with 10 to 0.01 nm wavelength [4]. X-ray examination is done by segmentation of bone region by wrenching out of bone-outline and spotting of fracture from the X-ray [5]. Even though X-ray can give excellent outcomes in usual fractures, various tissue injury cannot be exactly recognized. Another technique is the ultrasound imaging which is frequently used because of its nonionizing feature and cheapness [6]. But it requires medical staff with good expertise to clarify the images. Nowadays, X-ray data are digitized and interrupted on computers, known as computed tomography (CT) [7]. Intelligent bone fracture detection system is an intuitive categorizing system which is competent of detecting and categorizing bone fracture. In this method, the bone fracture location and shapes can be detected using different image processing technique. Finally, the classification phase is carried out by training back-propagation neural network and then tested on processed image [8].

For human health interpretation, new technologies that have emerged can provide efficient system. The technology was introduced for the identification of femur bone and the technique is named as gray level co-occurrence matrix [9]. Fracture surgery with help of computer is also an efficient growing system based on computer-integrated system which can be less exposed to radiations, but it requires preoperative CT study. Fracture of medial and anterior fossa can be recognized using CT scans [10]. But CT scan cannot identify minute and subtle fractures [4, 7].

Open fractures are easy to identify and can be identified faster, but hairline fractures take longer time to identify. Hairline fractures are very minute and subtle cracks in foots due to stress level or it may be due to any sport injury. It is not visible to our naked eye. These fractures may not occur parallel to the bone but it will always occur with an inclination [1]. These hairline fractures are very difficult to identify by the physician, as it is very subtle. Commonly, these fractures are diagnosed in people who continuously stand or walk in their daily life. This will create stress in bones which may lead to hairline fractures. As these are the common in our life, these types of fractures are life style disease. In mandibular hairline fracture detection, the hairline fractures are modeled as less cut in suitable weighted flow network. Here, the flow network can be constructed based on mandible geometry and initial

information [11]. Another detection technique for hairline fracture is Markova random field and Gibbs sampling [12]. The Wavelet packet Transform (WPT) is used for split the image into low frequency information, high frequency information, low-high frequency information, and high-low frequency information. The WPT is effectively used to denoise the image [13]. The specialty of WPT is that it provides both spatial and frequency domain characteristics of the original image. The X-ray and CT scan may not give better results regarding this hairline fractures.

In this chapter, we proposed a novel method, that is, detection of hairline fracture based on Wavelet Packet Transform and Canny edge detection. The Canny edge detection technique is used to detect the edges of the digital X-ray images. Compared with other edge detection techniques it will be give good result, as it will detect all the edges including the hairline fractures edges.

68.2 Materials and Methods

The proposed method detects the hairline fractures that are difficult to identify by the physician using the following procedures (Fig. 68.1).

The block diagram shows the process of detecting hairline fracture using WPT and Canny edge detection method. Digital X-ray image is taken as the input and it gets processed. Preprocessing includes Wavelet Packet Transform, wherein the original image is divided into low frequency and high frequency information. Then Canny operator is used to detect the edges. Using Canny detection procedure, the fracture is detected.

The X-ray digital image of the foot is taken as the input image. It is a ‘.png’ image file. Since it is gray scale image there is no need for image conversion. The output obtained after Canny detection method is shown in the figure. Canny is an edge detection operator. There are basically five steps in Canny detection.

- Smoothens the image using Gaussian filter to remove noise.
- Gradient magnitude and angle images are calculated.
- Apply nonmaxima suppression.
- Use double thresholding and connectivity analysis to detect and link edges.

$$fs(x, y) = f(x, y) * g(x, y) \quad (68.1)$$

$$g(x, y) = \exp((-x^2 + y^2)/2\sigma^2) \quad (68.2)$$

Fig. 68.1 Block diagram of hairline fracture detection



$$m(x, y) = \sqrt{(g_x^2 + g_y^2)} \tag{68.3}$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{g_x}{g_y}\right) \tag{68.4}$$

Where $f(x, y)$ is the input image, $g(x, y)$ is the Guassain filter, and $fs(x, y)$ is the filtered image. g_x and g_y represent the gradient in x and y directions respectively. $m(x, y)$ and $\alpha(x, y)$ represent magnitude and angle respectively.

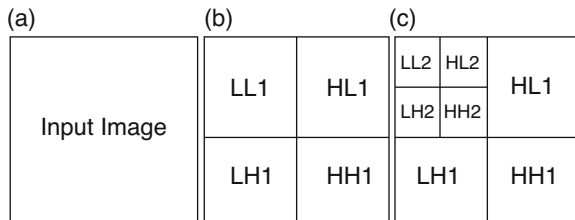
Sobel is an edge-detecting operator. It is used in computer vision, image processing particularly for edge detection. In this operator filters, having some integer values in horizontal and vertical directions are used to convolve the image, that is it uses two mask for identifying horizontal and vertical edges [14]. Thus, it calculates gradient in both horizontal and vertical direction. The mask values of Sobel operator are

$G_x =$	-1	0	1
	-2	0	2
	-1	0	1

$G_y =$	1	2	1
	0	0	0
	-1	-2	-1

The Wavelet Packet Transform plays a major role in image denoising. The basic idea of 2-D Wavelet Packet Decomposition is shown in the Fig 68.2. WPT is extended forms of Wavelet Transform [15]. By applying WPT on to an image, it splits into four coefficients LL1, HH1, LH1, and HL1, which are the 1-level image packets. The different image packets contained lowest frequency information, highest frequency information, and two packets of intermediate lowest and highest frequency information.

Fig. 68.2 2D-Wavelet packet transform decomposition (a) Original image (b) 1-level image packets (c) 2-level image packets.



68.3 Results and Discussions

The original X-ray digital image of the foot is undergoing preprocessing to suppress the high frequency and low frequency noise. The WPT decomposition is allotted to select the base for a particular type of noise denominated packets in the image. When in the deconstruction of the denoised image from the packets of the image using the WPT, the selected noise affected images packets are eliminated. Here we used 4-level 2D WPT decomposition for denoised the image. So, the deconstructed image is the denoised CT image.

Then denoised image and Gaussian filter are convolved. Then the gradient in x and y direction is calculated and it is denoted as g_x and g_y . The magnitude is calculated by taking the square root of sum of the squares of gradient in x and y directions which is denoted as $m(x, y)$. Figure 68.3 is a '.png' format gray-scaled digital X-ray image. This image is used to show the results of the proposed algorithm. Figure 68.4 shows the color map of the Canny edge detected image.

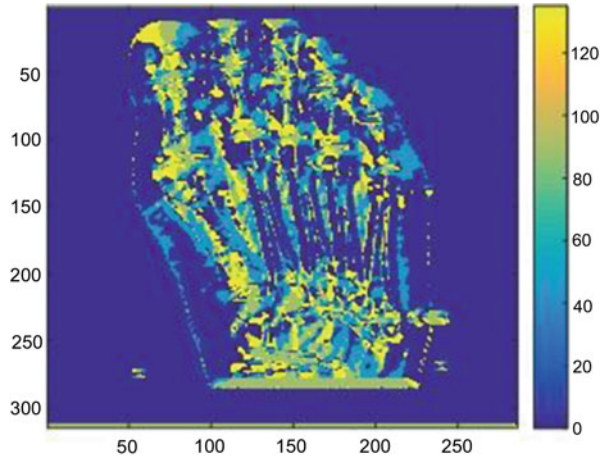
First-order derivative detect the thicker edges and second-order derivative detect the thinner edges. $m(x, y)$ contains width ridges around local maxima, in order to thin those ridges we use nonmaxima suppression. This can be done in several ways. For example, in a 3×3 region we can define four orientations: horizontal, vertical, $+45$, and -45 . If the edge normal is in the range of direction from -22.5 to 22.5 or from -157.5 to 157.5 , we can call the edges as horizontal edge.

Let d_1 , d_2 , d_3 , and d_4 denotes the four basic edge directions for a 3×3 region—horizontal, -45 vertical, and $+45$ respectively. The nonmaxima suppression scheme for a 3×3 region centered at every point in (x, y) and in $\alpha(x, y)$ can be formulated as finding the direction d_k that is closest to $\alpha(x, y)$. If the value of $m(x, y)$ is less than atleast one of its two neighbors along, d_k , let $gN(x, y) = 0$ (suppression) otherwise let

Fig. 68.3 Input Digital X-ray image



Fig. 68.4 Color map of the edge detected image



$gN(x,y) = m(x,y)$ where $gN(x,y)$ is the nonmaxima suppression image. Finally, false edge points are reduced by thresholding $gN(x,y)$. If we set the threshold as too low, false edges called false positives may occur, and if the threshold is set too high, then actual valid edge points will be eliminated (false -ve). This can be improved using Canny algorithm by using hysteresis thresholding.

It uses two thresholding – low threshold and high threshold. Canny suggested that the ratio of high to low threshold should be 2 to 1 or 3 to 1. The thresholding operation can be visualized as,

$$gNH(x,y) = gN(x,y) \geq \text{high threshold}$$

$$gNL(x,y) = gN(x,y) \geq \text{low threshold}$$

After thresholding $gNH(x,y)$ will be having fewer nonzero pixels than $gNL(x,y)$. But all the nonzero pixels in $gNH(x,y)$ will be contained in $gNL(x,y)$. In order to include strong and weak edge points in the resultant, we eliminate from $gNL(x,y)$, all the nonzero pixels from $gNH(x,y)$ by letting $gNL(x,y) = gNL(x,y) - gNH(x,y)$. After the thresholding operations, all strong pixels of $gNH(x,y)$ are assumed to be valid edge pixels and the mixed valid edge pixels are obtained by locating the next unvisited edge pixel in $gNH(x,y)$. Then mark as valid edge pixels, all the weak pixels in $gNL(x,y)$ that are connected to peak using eight connectivities. If all nonzero pixels in $gNH(x,y)$ has been visited then again locate the unvisited edge pixel in $gNH(x,y)$ and set to zero all pixels in $gNL(x,y)$ that were not marked as valid edge pixels. Figure 68.5 shows the output obtained after the Sobel operation. Figure 68.6 shows edge detected output using the Canny operator. It is clear from the results that more edges are detected using the Canny operator.

It can obtain the precise fracture area by subtracting the Canny detected output without fracture from Canny detected output with fracture. Thus, the final result contains only the expected fractured area which helps the physician to detect the hairline fracture without any confusion.

Fig. 68.5 Edge detection of the input image using Sobel operator



Fig. 68.6 Edge detection of the input image using Canny operator



68.4 Conclusion

Aiming for the problem of detecting very minute cracks in the foot due to stress or any sport-injury mainly known as hairline fractures, which are not visible to our naked eye. Since hairline fractures are subtle cracks, it is difficult to identify by the physician in its initial stages. This chapter proposed a new detection method of hairline fracture based on Wavelet Packet Transform and Canny Operator. The

4-level Wavelet Packet Decomposition is used to denoise the digital X-ray image. Also the chapter enhances the comparison between Canny and Sobel detection methods. Canny has the advantage of reducing the noise and keeping the fine image edges. The presence of hairline fracture may not be appear on digital X-ray, the results show that proposed method is capable of detecting hairline fracture in its initial stage itself.

References

1. Swischuk LE, Hernandez JA (2004) Frequently missed fractures in children (value of comparative views). *Emerg Radiol* 11(1):22–28
2. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J (2015) *Harrison's principles of internal medicine*, 19th edn. USA
3. John SD, Moorthy CS, Swischuk LE (1997) Expanding the concept of the toddler's fracture. *Radiographics* 17(2):367–376
4. Das G (2013) Bat algorithm based softcomputing approach to perceive hairline bone fracture in medical X-ray images. *Int J Comput Sci Eng Technol (IJCSSET)* 4(04):432–436
5. Bandyopadhyay O, Biswas A, Bhattacharya BB (2016) Long-bone fracture detection in digital X-ray images based on digital-geometric techniques. *Comput Methods Prog Biomed* 123:2–14
6. Nascimento L, Ruano MG (2015) Computer-aided bone fracture identification based on ultrasound images. In: 4th Portuguese Meeting on Bioengineering (ENBENG). IEEE, Porto, pp 1–6
7. Jacobsen C, Bech BH, Lynnerup N (2009) A comparative study of cranial, blunt trauma fractures as seen at medicolegal autopsy and by computed tomography. *BMC Med Imaging* 9 (18):1–9
8. Dimililer K (2017) IBFDS: intelligent bone fracture detection system. *Proc Comput Sci* 120:260–267
9. Chai HY, Wee LK, Swee TT, Salleh SH, Ariff AK (2011) Gray-level co-occurrence matrix bone fracture detection. *Am J Appl Sci* 8(1):7–16
10. Lee S, Lee JW, Jeong JW, Yoo DS, Kim S (2008) A preliminary study on discrimination of osteoporotic fractured group from nonfractured group using support vector machine. In: 30th annual international conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Vancouver, pp 474–477
11. Chowdhury AS, Bhandarkar SM, Robinson RW, Jack CY, Liu T (2011) Detection of hairline mandibular fracture using max-flow min-cut and Kolmogorov-Smirnov distance. In: International symposium on biomedical imaging: from nano to macro. IEEE, Chicago, pp 1962–1965
12. Chowdhury AS, Bhattacharya A, Bhandarkar SM, Datta GS, Yu JC, Figueroa R (2007) Hairline fracture detection using MRF and Gibbs sampling. In: Workshop on applications of computer vision (WACV). IEEE, Texas, p 56
13. Shui PL, Zhou ZF, Li JX (2007) Image denoising algorithm via best wavelet packet base using Wiener cost function. *IET Image Process* 1(3):311–318
14. Lo SCB, Liu CC, Freedman MT, Mun SK, Kula J, Lasser ME, Wang YJ (2009) PE-CMOS-based C-mode ultrasound: signal acquisition and time gating. In: *Medical imaging 2009: ultrasonic imaging and signal processing*. International Society for Optics and Photonics, Bellingham, p 726508
15. Biju KS, Jibukumar MG, Rajasekharan C (2018) EEG analysis using wavelet packet transforms on mean energy and mean Teager energy with artificial neuro-fuzzy system. In: Aloui F, Dincer I (eds) *Exergy for a better environment and improved sustainability 2*, Green energy and technology, vol 2. Springer, Cham, pp 627–635

Chapter 69

Image Encryption-Then-Compression System for Secure Transmission via Hybrid Henon Chaotic Map



P. Sridevi and J. Suguna

Abstract In recent years, the environments like military, government, medical field, cloud computing, and social networks deal with a large number of confidential images transmitted over the Internet. Therefore, it is very important to protect the image from unauthorized access during the transmission in an open network. Encryption is the most convenient technique to guarantee the security of images over public networks. To maximize the network utilization, the compression technique is used to reduce the size of the image by the channel provider who is having plenty of computational resources. This paper proposes an image Encryption-Then-Compression (ETC) system. Hybrid Chaotic method is used to encrypt the image, where Arnold map is used for confusion and Henon map is used for diffusion. Asymmetric Numerical Method (ANM) is used to compress the encrypted image. The experiment result shows that the proposed system is better in terms of compression performance, security, and computation time.

Keywords Image compression · Decompression · Hybrid chaotic encryption and asymmetric numerical method

Abbreviations

AC	Arithmetic Coder
AES	Advanced encryption standard
ANM	Asymmetric numerical method
CTE	Compression-then-encryption
ECC	Elliptical curve cryptography
ETC	Encryption-then-compression
HC	Huffman Coder

P. Sridevi (✉) · J. Suguna

Department of Computer Science, Vellalar College for Women, Erode, Tamil Nadu, India
e-mail: sridevi@vcw.ac.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_69

NPCR	Number of pixel change rate
ROI	Region of interest
UACI	Unified average changing intensity

69.1 Introduction

The development of information technology has a broad impact on the human ways of communication and digital data, particularly high-quality still images. In multimedia data, especially images have been increasing every day. Storage and transmission of images are not easy because they need high memory devices and high bandwidth network systems. Security of image also plays a vital role in this digital world. Encryption is the most effective technique to achieve image security and ensures that transmitted image is reliable. Compression and encryption are the two different techniques used in image processing to reduce the image file size and to provide security for images. This research mainly discusses an image Encryption-Then-Compression (ETC) system to enhance the efficiency and security of images during transmission. Symmetric and asymmetric encryption techniques are the two types of encryption. Mostly, combination of Encryption-Then-Compression technique uses symmetric encryption and lossless compression method. It shows that the ETC system focuses more on image security than the file size reduction. The proposed method uses Hybrid chaotic encryption to enhance the security. The chaos encryption schemes are composed of two steps: confusion and diffusion. Most of the encryption schemes faced some problems like lack of robustness and security. The key space is increased by choosing a high dimensional chaotic system. Repeated permutations are avoided but pixel values are changed by the diffusion function. Two of the common image compression techniques are lossless compression and lossy compression. The lossless compression technique produces no loss in the quality of image. Lossy compression technique is one which produces a minor loss of quality to the output image and this loss is almost invisible. The proposed ETC system comprises a hybrid chaotic method for encryption and Asymmetric Numeral Method (ANM) for lossless compression. A significant compression ratio can be achieved when compression is performed after encryption for both lossless and lossy compression.

The rest of this chapter is organized as follows. Section 69.2 explains the related researches briefly, and Sect. 69.3 describes the proposed methodology. Section 69.4 provides the experimental results and their discussions. Section 69.5 concludes the research work.

69.2 Related Works

K. Sakthidasan and Santhosh Krishna [7] proposed a method for encryption which consists of two stages known as confusion and diffusion stage. Complex chaotic maps are used to increase the complexity level of the algorithm. The 3D chaotic systems are used for permutation of the pixel position of plain image. Separate secret keys are used in the permutation and diffusion stage of the encryption process to increase the security level of an algorithm. Security analysis is carried out and simulation results showed that the encryption and decryption algorithm has good security and robustness. Zhou et al. [10] proposed a method of Compression-Then-Encryption (CTE) scheme based on hyper-chaos system and 2D sensing to overcome security risks. Cycle shift operation can change the values of the pixels efficiently. Simulation results showed that the Compression-Then-Encryption (CTE) scheme is effective, robust with a good compression performance and less security. J. Rui liu et al. [6] proposed a new algorithm for color image encryption using chaotic map and spatial bit-level permutation and showed that the algorithm was suitable for multimedia applications and real-time applications such as mobile phone services to achieve good encryption and low time complexity. A. P Shaikh and V. kaul [8] presented a hybrid model with the combination of symmetric algorithms such as Blowfish and Advance Encryption standard (AES). This cryptography maintained the data confidentiality and integrity. Ahmad et al. [2] proposed a new image encryption scheme based on chaotic maps and orthogonal matrices to perform encryption with higher security, it also supports partial encryption for a faster process. The experiment and security analysis showed that the encryption scheme is relatively secured, robust from channel noise, and output quality of a decrypted image is fairly good. K. Gupta et al. [4] proposed an ethical approach of block-based image encryption using chaotic map and proved the advantage of large key space, sensitivity to the key and resistance for brute force attack. Abdullah et al. [1] proposed an algorithm for image encryption which has three stages, namely, confusion, shuffling, and diffusion. The security measurements demonstrated that the joint weaknesses which existed in other encryption algorithms can be defeated in hybrid chaotic algorithm. Divija Ameta and Sandeep Upadhyay [3] proposed a hybrid approach for image encryption, which used different number of iterations in Elliptical Curve Cryptography (ECC) and Advance Encryption Standard (AES) techniques. The work not only focused on the normal encryption but also considered horizontal and vertical components along with region of interest (ROI) and provided more accuracy.

69.3 Proposed Methodology

The proposed system is Encryption-Then-Compression [ETC] system where hybrid chaotic encryption is used to encrypt the image and Asymmetric Numeral Method [ANM] is used to compress the encrypted image. The existing Compression-Then-Encryption (CTE) system provides less security for images and the encryption algorithm may remove the compressed bits in the image. To get an efficient system, the order of applying the compression and encryption needs to be reversed to meet all the requirements.

69.3.1 Image Encryption

The encryption algorithm consists of three stages. In first stage (confusion stage), the image pixels are confused, based on Arnold Cat map, where the pixel positions are shuffled over the whole image without changing the value of the pixels in the second stage and the image becomes unidentifiable. In third stage (diffusion stage), shuffled image is diffused by a key image which is generated by Henon map.

69.3.1.1 Confusion

First stage of encryption is confusion where pixels of the image are confused based on Arnold Cat map. The position of the pixels is shuffled over the whole image, but the value of the pixels remains the same. Therefore, the initial conditions and control parameters of Arnold Cat map serve as the first secret key. The significant property of Arnold cat map is that the position of image pixels are rearranged, and then it returns to the original pixel position after several iterations and produces the original image. The Arnold chaotic map is given by

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & p \\ q & pq + 1 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix} \text{mod } N$$

Here p and q represents control parameters, $(N*N)$ is the image size, (x_n, y_n) is the pixel location of the original image and (x_{n+1}, y_{n+1}) is the new pixel location in the shuffled image after applying Arnold transformation.

69.3.1.2 Shuffling

Pixel shuffling is the second stage of encryption. It is used to add more randomness and to increase the efficiency of the encryption. The shuffling algorithm represents

the second secret key. Initially, the image is separated into four blocks. Then each block is shuffled in predefined order, this process is repeated until it reaches the last quad. Finally, each entire quad is considered as single cell and shuffled in predefined order.

69.3.1.3 Diffusion

Third stage of encryption is diffusion which is used to improve the security. Diffusion aims to generate key image. Diffusion is based on Henon maps. It is a discrete-time dynamical system. It is one of the most studied examples of dynamical systems that exhibit chaotic behavior. The Henon map takes a point (x_n, y_n) in the plane and maps it to a new point.

$$\begin{cases} x_{n+1} = 1 - ax_n^2 + y_n \\ y_{n+1} = bx_n \end{cases}$$

The map depends on two parameters, namely, a and b , which for the classical Henon map have values of $a = 1.4$ and $b = 0.3$. Henon map is chaotic for the classical values. For other values of a and b , the map may be chaotic, intermittent or converge to a periodic orbit.

69.3.2 Compression using Asymmetric Numeral Method

Asymmetric Numeral Method (ANM) is used in this work to improve the performance of compression, ANM is an entropy coding method and lossless image compression method. Huffman Coder (HC) and Arithmetic Coder (AC) are the other entropy coder. Huffman Coder is faster but uses approximate probabilities with powers of two which leads to relatively low compression rates. The arithmetic coder uses nearly exact probabilities with larger computational time. ANM has the advantage to combine the compression ratio of arithmetic coder with the speed of Huffman Coder and it is used in image compression because of high performance and efficiency [5, 9].

69.4 Experimental Analysis

This research attempts to prove that the encryption scheme improves the security and compression performance when it is used with the ETC scheme. Experimental results reveal that the proposed hybrid Henon map encryption improves the security and compression performance. The proposed algorithm is analyzed based on various

Table 69.1 Compression performance of Colpitts–Henon chaotic system

Encryption	Compression	Images				
		Lena	Satellite	Medical	Baboon	Pepper
Hybrid Colpitts– Henon	ANM	147,090	198,464	199,444	156,925	153,091
	HC	144,030	171,433	182,435	154,453	152,354
	AC	142,694	195,029	160,260	144,054	152,238
Hybrid Colpitts– Duffing	ANM	146,933	198,348	199,301	156,724	152,945
	HC	143,876	171,319	182,363	154,289	152,192
	AC	142,521	194,901	160,159	143,823	152,111
Colpitts– Henon	ANM	146,867	198,277	199,213	156,647	152,847
	HC	143,789	171,234	182,267	154,231	152,142
	AC	142,456	194,845	160,089	143,756	152,035
Colpitts– Duffing	ANM	146,742	188,342	197,303	153,729	152,713
	HC	143,851	171,294	182,346	152,323	152,208
	AC	142,517	184,899	160,156	143,815	152,097
Predictive error method	ANM	134,267	183,277	194,213	135,647	125,647
	HC	133,789	181,234	192,267	135,231	125,142
	AC	132,456	179,845	190,089	134,756	125,035

performance measures like compression performance, computation time, NPCR, and UACI. Table 69.1 shows the compression performance of Hybrid Colpitts–Henon Chaosmap method in Bytes with three different compression techniques, namely, Asymmetric Numeral Method [ANM], Huffman Coder (HC), and Arithmetic Coder (AC). It is observed that the ANM compression with Colpitts–Henon Chaotic System gives good performance than the traditional Huffman and Arithmetic coders.

69.4.1 Number of Pixel Change Rate (NPCR)

Number of Pixel Change Rate (NPCR) is measured for the images to analyze the security of the ciphers. Table 69.2 shows the average of NPCR values and indicates that the sensitivity of the encrypted ciphers of 512×512 image is 98.18%, leading to the conclusion that the Proposed Hybrid Colpitts–Henon Chaotic System provides better encryption and it is very sensitive with respect to small pixel changes.

$$NPCR = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N D(i,j)$$

Table 69.2 Comparison of NPCR, UACI, and TIME

Encryption	Parameters	Images				
		Lena	Satellite	Medical	Baboon	Pepper
Hybrid Colpitts–Henon	NPCR	100	97.5	95.6	100	97.8
	UACI	33.2	31.9	38.1	36.9	35.1
	Time	1.212	2.421	1.759	2.228	1.924
Hybrid Colpitts–Duffing	NPCR	100	94.2	93.8	99.2	95.8
	UACI	36.6	33.7	37.1	37.5	29.2
	Time	1.468	1.979	2.312	2.429	2.222
Colpitts–Henon	NPCR	99	93.5	92.6	98.7	94.8
	UACI	34.4	33.5	35.4	37.8	34.2
	Time	2.134	2.743	2.676	2.864	2.684
Colpitts–Duffing	NPCR	98.4	95.3	97.4	97.5	96.4
	UACI	35.4	35.6	36.7	33.5	34.7
	Time	2.775	2.065	2.883	2.556	2.948
Predictive error method	NPCR	91.6	91.6	94.6	97.8	92.5
	UACI	33.2	35	36.2	31.2	30.4
	Time	2.267	2.343	2.912	2.865	2.954

$$D(i, j) = \begin{cases} 0, & \text{if } Im\ o(i, j) = Im\ c(i, j) \\ 1, & \text{if } Im\ o(i, j) \neq Im\ c(i, j) \end{cases}$$

69.4.2 Unified Average Changing Intensity (UACI)

UACI determines the average change in pixel intensities in corresponding positions in the two encrypted images as a percentage with F , the maximum supported pixel value ($F = 255$ in gray scale images). From the Table 69.2, the average value of UAIC is 35.04%. UACI value of the proposed hybrid Colpitts–Henon encryption scheme is higher than other encryption schemes. The higher value indicates the rate of one pixel change which shows that the proposed Hybrid Colpitts–Henon Chaotic System provides better encryption.

$$UACI = \frac{100}{MN} \sum_{i=1}^M \sum_{j=1}^N \frac{|Im\ o(i, j) - Im\ c(i, j)|}{255}$$

69.4.3 Computation Time

In Table 69.2, the computation time of encryption schemes are given. It is found that the average computation time is stable for all the encryption schemes. From the analysis, it is clear that the proposed hybrid Colpitts–Henon chaotic based encryption system with ANM compression shows better result in ETC system.

69.5 Conclusion

Image encryption and compression are an extremely important part of modern computing. This work designed an efficient image Encryption-Then-Compression (ETC) system and proposed a new hybrid chaotic encryption for secure image transmission and storage. In this ETC system, Hybrid Chaotic map algorithm significantly increases the security and the efficiency of encryption. The encrypted image is compressed using Asymmetric Numeral Method (ANM). The experiment result shows that the proposed system is better in terms of compression performance, security, and computation time.

References

1. Abdullah HN, Hamsa A (2017) Image encryption using hybrid chaotic map. In Current research in computer science and information technology (ICCCIT). 2017 International conference on IEEE, April, pp 121–125
2. Ahmad J, Khan MA, Hwang SO, Khan JS (2016) A compression sensing and noise-tolerant image encryption scheme based on chaotic maps and orthogonal matrices. *Neural Comput Appl* 28(1):953–967
3. Ameta D, Upadhyay S (2017) A hybrid approach for image encryption using different number iterations in ECC and AES techniques. *Int J Comp Appl* (0975 – 8887) 175(3):10–12
4. Duda J (2014) Asymmetric numeral Systems: entropy coding combining speed of Huffman coding with compression rate of arithmetic coding, arXiv:1311.2540V2[CS.IT]6, Jan 2014
5. Gupta K, Gupta R, Agrawal R, Khan S (2015) An ethical approach of block based image encryption using chaotic map. *Int J Secur Appl* 9(9):105–122
6. Rui Liu, X Tian (2012) New algorithm for color image encryption using chaotic map and spatial bit level permutation. *J Theor Appl Inf Technol* 43(1) © 2005–2012 JATIT & LLS
7. Sankaran KS, Krishna BVS (2011) A new chaotic algorithm for image encryption and decryption of digital color images. *Int J Inf Educ Technol* 1(2):137–141
8. Shaikh AP, kaul V (2014) Enhanced security algorithm using hybrid encryption and ECC. *IOSR J Comp Eng (IOSRJCE)* 6(3):80–85
9. Sridevi P, Suguna J (2017) An efficient encryption then compression system using asymmetric numeral method. *IJET* 9(5):3680–3688
10. Zhou N, Pan S, Cheng S, Zhou Z (2016) Image compression–encryption scheme based on hyper-chaotic system and 2D compressive sensing. *Opt Laser Technol* 82:121–133

Chapter 70

Analysis of Primary Emulsion Attack in Cognitive Radio Using Distributed On-Demand Routing Protocol



Neelaveni Rangaraj and Sridevi Balu

Abstract The aim of this chapter is to design a novel framework for cognitive radio to overcome the security-based challenge by considering authentication and confidentiality. Particularly, the chapter focuses on the primary emulation attack, as it gives the authentication for the unlicensed user to use the unused the spectrum. These unlicensed users are considered as the secondary users and to authorize to use the spectrum only for the required period without compromising the security of the primary user. The distributed on-demand routing protocol is used in cognitive radio, and hence it can be used for the group of users sharing the same spectrum. RSA with the distributed on-demand routing protocol yields a secure key for sharing that particular session within the users. A comparison between the classical protocols for generating the secret key with Diffie–Hellman algorithm and other protocols is also done in this work by analyzing their vulnerabilities.

Keywords Distributed on-demand · RSA algorithm · Authentication · Spectrum sharing

Abbreviations

ACK	Acknowledgment
CBS	Cognitive Base Station
CR	Cognitive Radio
CRV	Credit Risk Value
CS	Cognitive Sensing
DORP	Distributed On-Demand Routing Protocol
DSDV	Destination Sequenced Distance Vector

N. Rangaraj (✉)

MNM Jain Engineering college, Chennai, Tamil Nadu, India

S. Balu

Velammal Institute of Technology, Chennai, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_70

PU	Primary User
PUE	Primary User Emulsion
RSA	Rivest Shamir and Adleman (Public Key Encryption Technology)
ZRP	Zone routing protocol

70.1 Introduction

Cognitive Radio can change its transmitter parameter based on interaction with the environment in which it operates. It comprises of licensed user as well as unlicensed user. Its characteristics are spectrum sensing, spectrum analysis, and spectrum decision. The spectrum-sensing techniques in cognitive radio transmitter include cyclo-stationary-based detection, energy detection, and matched filter detection. Matched filter detection sensing is performed by correlating the observed signal with the known sample to detect the presence of primary users. Sensing the energy of the signal with the estimated and the limited value is the energy detection method. The limited value indicates the channel can be used by the user. The drawback is that it cannot distinguish between signals of the primary user from those of the secondary user. Cyclo-stationary-based detection is used for detecting periodicity of the received signal. One of the sensing techniques is the cooperative sensing technique which is further classified as centralized and decentralized spectrum technique in which former is a cluster head or server that collects the sensing information and transmits this information to control the cognitive radio traffic. The distributed sensing is in which cognitive nodes inform the other nodes which are in the same group to choose the spectrum. Hybrid sensing is very similar to distributed system sensing. In this technique, when the primary user arrives, it vacates the channel immediately without informing other nodes. The primary user as a licensed user gives permission to access the spectrum for the secondary users. When the secondary user acts as a primary user using the identity of the primary user to know about the detailed information about the spectrum, then it is said to be the primary emulsion attack.

70.2 Previous Works

Mandakini Gupta et al. [1] described primary user emulsion (PUE) attack as one of the most important threats of spectrum sensing for wireless cognitive radio network. Based on the wireless signal, a PU is detected in a given band, and all secondary users should avoid accessing that band. However, when a secondary user is detected, other secondary users may choose to share that same band. In a primary user attack, a malicious secondary user tries to gain priority over the secondary user by transmitting signals that emulate the characteristics of primary user. In this chapter, the radio

software of a cognitive radio is modified to change its emulsion characteristics (i.e., modulation, frequency, power, etc.) so that the emulsion characteristics resemble those of a PU.

R. Akhila et al. [2] analyzed that in a wireless ad hoc network, the spectrum will not be sufficiently used. The licensed user can use the licensed spectrum who is said to be the primary user, whereas the unlicensed user can use the spectrum when the licensed users are not using it. This authentication to the unlicensed user must be given in a proper manner to maintain the secrecy of the data. The unlicensed secondary users may transmit and restore the fake information about the spectrum for its future use, thus causing a danger to the message spread through the spectrum to the primary user. Those secondary users will be known as the selfish users. They will degrade the network performance. Here, a method which uses credit risk value (CRV) is being proposed to identify the selfish users to improve the network performance and efficiency.

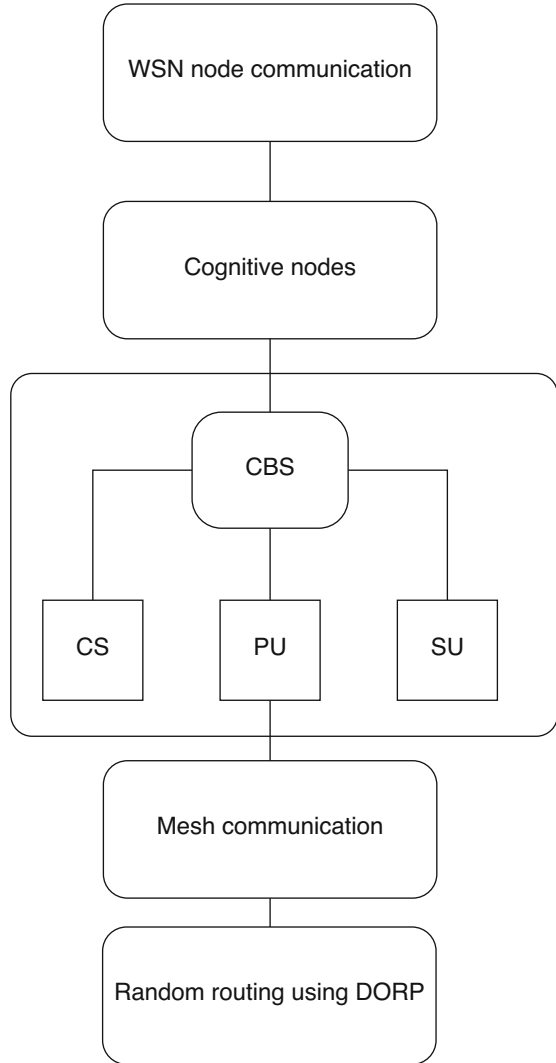
Rohit Chakravarthy et al. [3] also described primary user emulation (PUE) attack as one of the major concerns and problems in cognitive radio networks. In this chapter, a primary use authentication scheme is proposed to solve the PUE attack problem by employing an underlay RF fingerprint in the primary user signal. Specifically, an underlay waveform is introduced on the top of the header of the legitimate primary user signal. As a result, the underlay waveform exhibits a unique and different cyclo-stationary feature than the primary user signal. The PUE attack signal, on the other hand, will not reveal this cyclo-stationary feature and fail the authentication. Throughput Maximization in Cognitive radio networks [4], Security enhancement using key agreement protocols [5], analysis of the sensing techniques in CR network [6], and the Primary emulation attack rectification by physical layer network coding [7] are explored.

70.3 Proposed Work

This chapter focuses to enhance the security of primary user during the primary emulsion attack in a cognitive radio network. Cognitive radio (CR) is the enabling technology for supporting dynamic spectrum access. The functionality of cognitive radio includes the characteristics of power, modulation techniques, to find its geographic location using the transceiver to form a group of nodes, sensing the spectrum for the primary user, implementing the key exchange methodology for a secured communication, and to make it applicable for mobile applications.

To represent these functions of CR, a group of nodes are created for representing the communication in CR. The cognitive base station (CBS) acts as a central medium for communication. CBS finds a cognitive sensing (CS) which is an optimum node for communication in each group of nodes as shown in Fig. 70.1. Each group comprises of primary and secondary users who in a real-time example can be differentiated by the cost as licensed and unlicensed users. When the primary user wants to communicate, the access is taking place through cognitive sensing. The RSA algorithm is used in the authentication process between a cognitive

Fig. 70.1 Basic block diagram



sensed node and a primary user. Similarly, between CS and CBS, the encryption and decryption of messages take place using this algorithm. When the primary user is not using the spectrum (spectrum underutilization), secondary users communicate via primary user by the same authentication process. Though CR has been proposed as a promising solution for improving spectrum utilization, the major drawback is the primary emulsion attack. In CR network, primary users have priority over secondary users when accessing the wireless channel. The primary emulsion attack is said to happen when a malicious secondary user exploits this spectrum access acting as a primary user. This has to be avoided to achieve a successful communication. Hence, we introduce a group-leader technique to secure the primary user.

70.4 Implementation

The implementation of the security enhancement of the cognitive radio is done with the help of network simulator.

70.4.1 Initial Setup

When the nodes are randomly deployed, the communication is carried out using the destination sequenced distance vector algorithm. This algorithm is similar to the distance vector algorithm in which the routing table is updated periodically, where each node informs about its neighboring node and there by forming the network topology. In DSDV, sequence number is used to avoid the formation of loop. Though this routing is carried out, the efficiency in communication is hindered by the primary emulsion attack, where the secondary users try to gain priority over the primary user by transmitting signals that emulate the characteristics of the primary users.

70.4.2 Distributed On-Demand Routing Protocol

In Distributed On-Demand Routing Protocol (DORP), the cumulative delay between the end nodes is calculated by determining the delay in each route. Hence, the best distributing path is found using this protocol. The added advantage of this routing technique is an improved range for communication as shown in Fig. 70.2.

70.4.3 RSA Algorithm

Key exchange between the CBS and the primary user plays the vital role in the security enhancement of the cognitive spectrum which prevents the primary emulsion attack. The public key generated by the RSA algorithm in the CBS is decrypted in the primary user using the private key. When there is a mismatch in the key between them then it is said to be an attack by the secondary user. If the key matches, then the authentication to use the spectrum can be given to the user who generated the key. The following steps are followed for the public key generation in CBS.

Step1: Select two prime numbers p and q .

Step2: Calculate $n = p * q$.

Step3: Calculate $\phi(n) = (p-1) * (q-1)$.

Step4: Select a variable e (for encryption) in which it should satisfy the condition $1 < e < \phi(n)$.

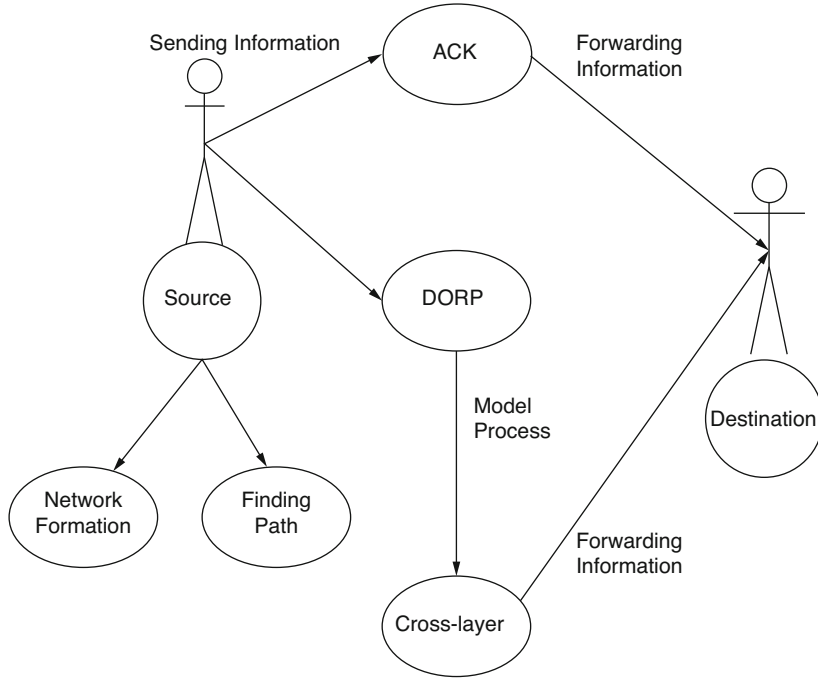


Fig. 70.2 Functional diagram of distributed on-demand routing protocol

- Step5: Calculate d such that $d \cdot e \pmod{\phi(n)} = 1$.
- Step6: For calculating private key $K_r = \{d, n\}$ is used.
- Step7: For calculating public key $K_u = \{e, n\}$ is used.
- Step8: For encryption the plain text M should be less than n .
- Step9: Cipher text C in CBS can be calculated by $C = M^e \pmod{n}$.
- Step 10: For decrypting the plain text $M = C^d \pmod{n}$.

70.5 Comparisons

The Diffie–Hellman algorithm is the better option for an insecure communication channel, even when there is no interchange of keys or messages in between the end users. However, the RSA algorithm is used for the channel to secure the user authentication as because of their symmetric key exchange methodology as a digital signature which is little faster than the Diffie–Hellman algorithm. Even though the RSA algorithm has various advantages, it can be used for applications like an emergency which considers without the possibility of usage of other spectrums. However, for military applications, the Diffie–Hellman algorithm can be used for asymmetric key exchange.

The mobile ad hoc protocols are classified into table-driven and on-demand-driven protocols. As the spectrum for the user will be selected only when needed, distributed on-demand protocols are used, even though the table can be maintained for the primary user, each time CBS need not store the routing loops as it may lead to insecure communication. The hybrid, zone routing protocol (ZRP), can be used as it is applicable for versatile environments. However, the significance of ZRP is the usage for wide range of area. As this chapter concerns only a short range, and keeping the parameters payload and latency, Distributed On-Demand Protocol is preferred over the Zone Routing Protocol (ZRP).

70.6 Conclusions

The purpose of this chapter is to give a new user some basic idea of how the simulator works, how to set up simulation networks, where to look for further information about network components in simulator codes, how to create new network components, etc., mainly by giving simple examples and brief explanations based on our experiences. Although all the usages of the simulator or possible network simulation setups may not be covered in this project, the project should help a new user to get started quickly.

References

1. Gupta M, Jain A, Soni A (2016) A survey: render of PUE attack in cognitive radio compressed by software defined radio. IEEE, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), pp 4029–4034,
2. Priyadharshini RA, Haimavathi KU (2016) Detection of attacks and Countermeasures in cognitive radio network. International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp 1102–1106
3. Chakravarthy R, Huang K, Zhang L, Wu Z (2017) Primary user authentication of cognitive radio network using underlay waveform. Cognitive communications for aerospace applications workshop, pp 1–5
4. Wang CL, Chen HW, Tsai ZY (2012) Throughput maximization for cognitive radio networks with wideband spectrum sensing. IEEE wireless communications and networking conference, pp 1293–1298
5. Vizvari S, Berangi R, Nematollahi K (2016) Authentication and authorizing scheme based on UMTS AKA protocol for cognitive radio network. IEEE KBEI, pp 118–123
6. Zheng J, Chen C-H, Cheng J-y, Shi L (2009) Cognitive radio: methods for the detection of free bands. IEEE Int Conf Netw Secur Wireless Commun Trust Comput 2:343–345
7. Gope J, Dutta P, Bhadra S, Das S, Jana S, Dalmia N, Choudhury S (2017) Analytical study of primary user emulsion attack detection techniques in cognitive radio ad hoc network. IEEE 8th annual ubiquitous computing, electronics and Mobile communication conference (UEMCON), October, pp 392–395

Chapter 71

Heart Disease Prediction Using Retinal Fundus Image



R. Rekha, V. P. Brintha, and P. Anushree

Abstract Heart disease increases the mortality rate in the recent years across the world. So it is necessary to develop a model to predict the heart disease occurrence as early as possible with higher rate of accuracy. In this study, the cardiovascular disease is predicted by non-invasive method with the retinal image data. In this system, the retinal fundus image data are used to predict the heart disease occurrence. Cardiovascular disease can be detected from the changes in microvasculature, which is imaged from retina. The prediction of a disease is by considering features like age, gender, smoking status, systolic blood pressure, diastolic blood pressure, and HbA1c. Risk factors for heart disease occurrence are detected from the microvasculature of segmented retinal fundus image using MATLAB. The main objective of the proposed system is to predict the occurrence of heart disease from retinal fundus image with higher rate of accuracy.

Keywords Cardiovascular disease · Retinal fundus image · Arteriole · Venule · Segmentation

Abbreviations

ANN	Artificial Neural Network
AVR	Arteries and Veins Ratio
CRAE	Central Retinal Artery Equivalent
CRVE	Central Retinal Vein Equivalent
CVD	Cardiovascular Disease
HD	Heart Disease
SVM	Support Vector Machine

R. Rekha (✉) · V. P. Brintha · P. Anushree
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

71.1 Introduction

Cardiovascular disease is one of the major causes to increase the mortality rate in the developed world. Therefore, the prediction of cardiovascular disease is very important and adequate to decrease the mortality rate. There are many techniques available to detect the occurrence of cardiovascular disease. Yet, they are costly to detect the disease and also take more time. The relation between heart and eye is high. Retina is one of the important features that help in direct microcirculation. Retina provides a window for detecting changes in microvasculature relating to the development of cardiovascular disease. Analysis of the retinal microvasculature provides information about the structure as the function of the vessels and this information is obtained easily. Structural changes in retina are one of early indications of the presence and severity of coronary artery disease. Research work has to be done before using them as clinical screening tool to estimate risk of developing a heart attack. Cardiovascular disease is defined as the problems caused in the blood vessels, circulatory system as well as the deformities in the heart. Numerous deaths occur due to heart disease in this world.

71.2 Related Work

The literature survey on heart disease prediction from retinal fundus image is shown in Table 71.1.

Table 71.1 Literature survey

Ref. no.	Problem Statement and Methodology	Merits
[1]	Neuro-genetic model by multi-objective genetic algorithm	Accuracy is 89.58%
[2]	Soft-computing technique with clinical parameters	Classifies patient
[3]	Decision tree combined with pre-pruning, post-pruning.	Features are stored
[4]	Non-linear proximal SVM for ischemic HD	Accuracy improved
[5]	DT, NB K-NN and NN for classification of HD	NB improves accuracy
[6]	MP, SVM, ANN, DT, K-NN, NN BP and GA at early stage	Accuracy comparison
[7]	Multilayer perceptron neural network	Complex problem
[8]	Threshold for segmentation of retinal image	Accuracy-pixels ratio
[9]	'Prewitt' and 'Sobel' operators for prediction, detection	Early detection
[10]	Retinal vascular calibre for CHD using AVR	Feature-Venule
[11]	CVD prediction using (A/V ratio) = CRAE/CRVE	Long-term risk factors
[12]	Stroke prediction using Dempster-Shafer theory	Accuracy is 61%
[13]	Ensemble classifiers with PSO for feature selection	Accuracy is 90.56%

71.3 Methodology

71.3.1 Pre-processing Techniques

The retinal image suffers from imperfections like poor contrast and noise, which need to be reduced or eliminated before processing. Pre-processing of the retinal fundus image is done using Gaussian filter. The image dataset is fed, and then it is pre-processed by subtracting original image with the Gaussian value of the image, to remove the Gaussian noise. First, the boundary is extracted before segmenting the blood vessel. The blood vessels of the retinal image are segmented. Flow diagram for the proposed methodology is shown in Fig. 71.1.

71.3.2 Segmentation Techniques

Line-tracking algorithm is used for segmenting the blood vessel of the retinal fundus image [14]. The tracking algorithm is performed after processing the scribble by Gabor filter-based algorithm to group scribble line strokes together, and therefore, information on the orientation of each line is known. The basis of the perceptual selection of the line-tracking direction at junction points based estimated in orientation. Since the proposed line-tracking algorithm is algorithmically independent of the Gabor filter grouping algorithm, any other line group orientation estimation method can be used to obtain the required orientation estimates [15].

71.3.3 Edge Detection Techniques

The major methods for detecting edges may be grouped into two categories:

- (i) Gradient
- (ii) Laplacian.

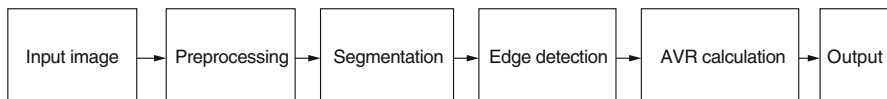


Fig. 71.1 Flow diagram for the proposed system

In gradient method, the edges are detected by looking for the maximum and minimum in the first derivative of the image, while in the Laplacian method, edges are detected by searching in the second derivative for the zero crossings of the image. The local changes of intensity in an image can be identified if the edge is more significant. Normally, edges occur on the boundary between two different regions in an image. From the edges of the image, important features can be extracted [16]. Sobel edge detection algorithm has the main advantage that it has smoothing effect on the random noise in the image due to average factor. In this edge of the segmented blood vessel is detected using the Sobel edge algorithm.

71.4 Implementation

During implementing of our method, the process is subdivided into four steps:

- (i) Pre-processing
- (ii) Segmentation
- (iii) Edge detection
- (iv) Calculation of A/V ratio

Retinal fundus image dataset is collected from Kaggle. About 1022 retinal fundus image are taken for processing.

71.4.1 Segmentation

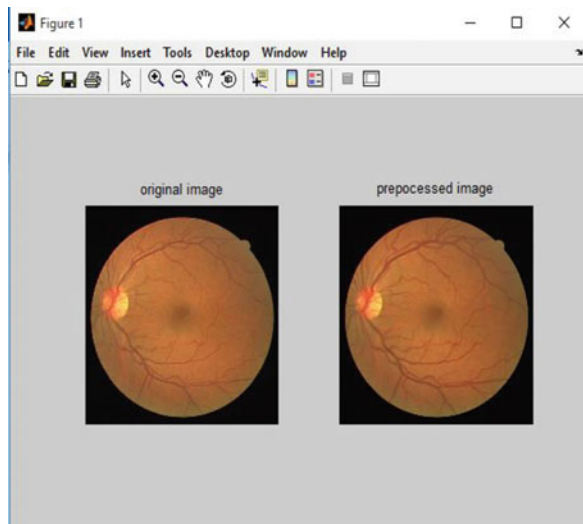
This dataset consists of retinal fundus images which were collected by EyePACS Eyepacs – healthcare foundation. The retinal fundus image is pre-processed before processing it. The original retinal fundus image, which differentiates arteriole and venules, is shown in Fig. 71.2. Arterioles are generally thicker than the venules. People with smaller AV Ratio is likely to have hypertension, and higher the AV Ratio indicates systolic and diastolic blood pressure. Largest retinal venular calibres have the highest incidence of stroke. The size of arterioles are associated with smoking status. People with larger sized arterioles have lower systolic but higher diastolic blood pressure and total cholesterol level; and tend to be current cigarette smoker [17]. Retinal fundus image after pre-processed image is shown in Fig. 71.3.

Line Tracking Algorithm The pixel area boundaries will be determined by the histogram. The blood vessels are traced by the threshold value corresponding to the frequency of the intensity image. Binary value is used to represent the pixel while tracking the image [18]. Segmented blood vessels of the retinal image after segmentation process are shown in Fig. 71.4. Processing of retinal fundus image with line tracking algorithm is shown in Fig. 71.6.

Fig. 71.2 Retinal fundus image



Fig. 71.3 Image after pre-processing



71.4.2 Edge Detection

After the segmentation of blood vessel from the retinal fundus image, boundaries of the segmented blood vessel are extracted. In order to find the vessel width of arteriole and venule, Sobel edge detection algorithm is used. Sobel edge detection extracts the boundary edge of the image. Then, vessel width is identified, according to the threshold level the heart disease is predicted for a given patient's retinal fundus image. Edge detected from segmented image by Sobel edge detection shown in Fig. 71.5.

Fig. 71.4 Image after segmentation

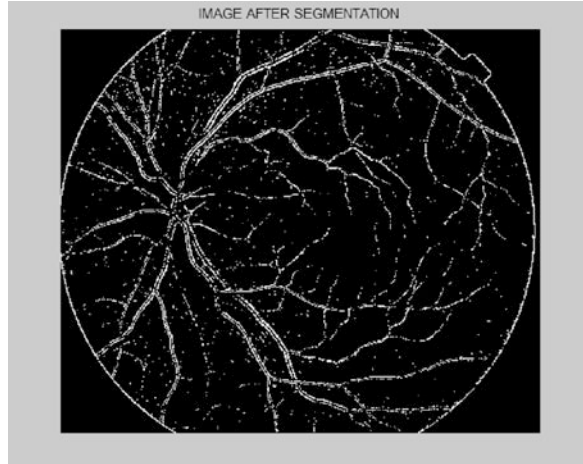
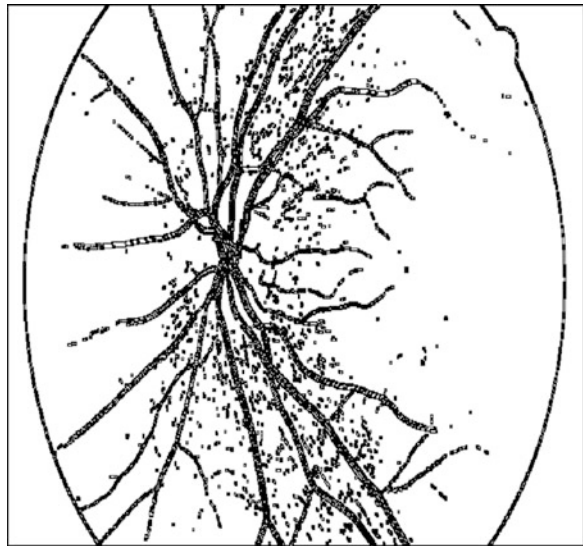


Fig. 71.5 Edge detection



71.4.3 *Calculating of A/V Ratio*

Changes in vessel structure have a major role in the disease prediction and diagnosis. The disease may occur due to atherosclerotic changes. Hypertension and atherosclerosis change the ratio between the retinal arteries and veins (AVR). Change in the A/V ratio is also associated with the increased risk of stroke and myocardial infarction. The central retinal artery equivalent (CRAE) and central retinal vein

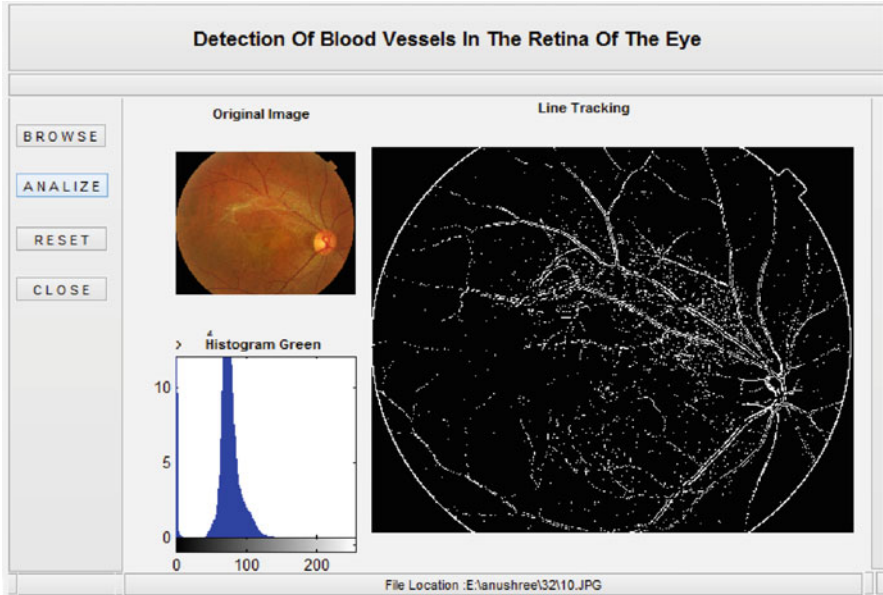


Fig. 71.6 Processing image during line tracking segmentation process

equivalent (CRVE) estimate the arteriole and the venular diameter inside Optic Disc (OD) region of vessel in retina (Fig. 71.6).

71.5 Result

Heart disease is predicted by measuring the A/V ratio. The threshold for predicting the heart disease is $A/V < 0.6$. The proposed method allows us to classify two classes, heart-disease person and healthy person with the data, after segmenting the retina fundus image. Prediction of heart disease using A/V ratio is shown in Table 71.2.

71.6 Conclusion

Prediction of heart disease occurrence for a patient by non-invasive method is very essential with the increase population and to reduce mortality rate in the developing world. Prediction of risk for the occurrence of cardiovascular disease from the given input of retinal dataset with high accuracy is developed in this study. Furthermore, this work can be extended by using live prediction with retinal fundus photography.

Table 71.2 Prediction of heart disease using A/V ratio

Arteriole	Venule	A/V Ratio	HD
0.05291	0.1389	0.0101	No
0.05295	0.11244	0.0140	No
0.03968	0.09260	0.1133	Yes
0.05291	0.13229	0.0105	No
0.05291	0.15875	0.0088	No
0.18520	0.07937	0.0617	Yes
0.13229	0.05291	0.0661	Yes
0.15875	0.05291	0.0793	Yes
0.13322	0.05291	0.0661	Yes
0.07937	0.00396	0.0529	No

References

- Murthy HN, Meenakshi M (2014) Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease. In (I4C), pp 329–332. IEEE
- Quijano M, Tello JP, Cadena A Heart disease predictor system based on artificial neural network. In (PAHCE), 2013 Pan American, pp 1–4. IEEE
- Mahmood AM, Kuppa MR Early detection of clinical parameters in heart disease by improved decision tree algorithm. In (VCON), 2010, pp 24–29. IEEE
- Soman KP, Shyam DM, Madhavdas P Efficient classification and analysis of ischemic heart disease using proximal support vector machines based decision trees. In TENCON 2003, vol. 1, pp 214–217. IEEE
- Peter TJ, Somasundaram K An empirical study on prediction of heart disease using classification data mining techniques. In (ICAESM), 2012, pp 514–518. IEEE
- Banu NS., Swamy S Prediction of heart disease at early stage using data mining and big data analytics: a survey. In (ICEECCOT), 2016, pp. 256–261. IEEE
- Sonawane JS, Patil DR Prediction of heart disease using multilayer perceptron neural network. In (ICICES), 2014, pp 1–6. IEEE
- Oliveira WS, Ren TI, Cavalcanti GD Retinal vessel segmentation using average of synthetic exact filters and hessian matrix. In (ICIP), pp 2017–2020. IEEE
- Panchal P, Bhojani R, Panchal T An algorithm for retinal feature extraction using hybrid approach. Proc Comp Sci 79:61–68
- Shylaja SS Algorithmic approach for prediction and early detection of diseases using retinal images. In CGIV'07 2007, pp 501–505. IEEE
- Seidelmann et al Retinal vessel calibers in predicting long-term cardiovascular outcomes: the atherosclerosis risk in communities study. Circulation 134(18):1328–1338
- Peñafiel et al Associating risks of getting strokes with data from health checkup records using Dempster-Shafer Theory. In (ICACT), 2018, pp 239–246. IEEE
- Yekkala I, Dixit S, Jabbar M A prediction of heart disease using ensemble learning and particle swarm optimization. In 2017, pp 691–698. IEEE
- Bartolo A, Camilleri KP, Fabri SG, Borg JC Line tracking algorithm for scribbled drawings. In ISCCSP 2008, pp 554–559. IEEE
- Sharma A, Rani S An automatic segmentation & detection of blood vessels and optic disc in retinal images. In (ICCSP), 2016, pp 1674–1678. IEEE
- Gupta S, Mazumdar SG Sobel edge detection algorithm. Int J Comp Sci Manage Res 2(2):1578–1583
- Poplin et al Predicting cardiovascular risk factors from retinal fundus photographs using deep learning. arXiv preprint arXiv:1708.09843
- <https://www.kaggle.com/c/diabetic-retinopathy-detection>

Chapter 72

Blind Speech Enhancement Using Adaptive Algorithms



P. Shanmuga Priya and S. Selva Nidhyananthan

Abstract Speech enhancement is used in all wireless telecommunication systems. Some of the interferences like white noise, periodic noises like human noise, and room reverberations may affect the speech quality. The delinquent of speech enhancement is communicated in this chapter. Least mean square (LMS), normalized least mean square (NLMS), and dual fast normalized least mean square (DFNLMS) are discussed in this chapter. The objective criteria are calculated for segmental signal-to-noise ratio, segmental mean square error, signal-to-noise ratio, and comparison results of the adaptive algorithms are discussed.

Keywords Blind Source Separation (BSS) · Least Mean Square (LMS) · Normalized Least Mean Square (NLMS) · Dual Fast Normalized Least Mean Square (DFNLMS)

Abbreviations

BSS	Blind source separation
DFNLMS	Dual fast normalized least mean square
LMS	Least mean square
NLMS	Normalized least mean square
SEGMSE	Segmental MSE
SEGSNR	Segmental SNR
SNR	Signal-to-noise ratio

P. Shanmuga Priya (✉) · S. Selva Nidhyananthan
Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India
e-mail: nidhyan@mepcoeng.ac.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_72

72.1 Introduction

The quality of speech and intelligibility of speech are significant for the transmission of communication. The objective is to improve the noisy speech quality by using various algorithms. With the help of the enhancement algorithm, noisy speech signals are reduced. The acoustic noise [1] taken from the microphone is undesirable, as it limits the perceived quality or intelligibility of the original speech signal. The researchers are working on effective reduction of noise. The signal corruption caused by noise is used to decrease using adaptive filters [3]. Noise cancellation technique is used to improve the signal-to-noise ratio [4]. The adaptive algorithms were suggested by the researchers. To reduce the noise and improve the speech signal [2], several adaptive algorithms were proposed [5]. Recursive least square (RLS) algorithms are discussed in [6], least mean square algorithm is proposed in [7] and affine projection algorithm in [8]. The forward BSS and backward BSS are discussed in [9, 10]. This chapter is organized as follows: in Sect. 72.2, we discuss about the adaptive algorithms for blind speech enhancement; in Sect. 72.3, the experimental results of the various adaptive algorithms are given; and in Sect. 72.4, the study conclusion is given.

72.2 Adaptive Algorithms for Blind Speech Enhancement

72.2.1 *Least Mean Square*

In various application, LMS algorithm is used because of its simplicity and reduced computational complexity. LMS algorithms are used to yield the mean square of the error signal, that is, difference between the desired and the original signal. It is based on the steepest descent method. The model of LMS filter is to obtain the convergence of the filtered weight by updating the filter weights. Each iteration of LMS involves three steps: filter output, estimation error, and tap weight adaptation.

72.2.2 *Normalized Least Mean Square*

NLMS filter is same as that of the least mean square. From first iteration to the next, the weights of an adaptive filter should be changed in a minimal order. The drawback of LMS is selection of step size. To overcome this difficulty we can move to the Normalized Least Mean Square (NLMS) algorithm. The advantages of NLMS filter to normalize the step size.

72.2.3 Dual Fast Normalized Least Mean Square (DFNLMS)

Forward BSS and backward BSS structures are generally used, for speech enhancement and acoustic noise reduction application. The forward BSS and backward BSS are combined with several adaptive algorithms. The output of the convolution mixture model is given by the following relations:

$$m_1(n) = s(n) + h_1(n) * b(n) \quad (72.1)$$

$$m_2(n) = b(n) + h_2(n) * s(n) \quad (72.2)$$

Where $*$ represents the convolution operation; and $s(n)$ is the original speech and $b(n)$ is the noisy speech signal. The output $u_1(n)$ and $u_2(n)$ of the proposed algorithm are given as,

$$u_1(n) = s(n) - w_1^T(n)M_2(n) \quad (72.3)$$

$$u_2(n) = s(n) - w_2^T(n)M_1(n) \quad (72.4)$$

where $M_1(n)$ and $M_2(n)$ are given in [10]. The update weights of the adaptive filter $w_1(n)$ and $w_2(n)$ are given by the relations,

$$w_1(n+1) = w_1(n) - \mu_1(u_1(n)C_1(n)) \quad (72.5)$$

$$w_2(n+1) = w_2(n) - \mu_2(u_2(n)C_2(n)) \quad (72.6)$$

Where $0 < \mu_1, \mu_2 < 2$ are the two step sizes control of convergence behavior of the cross-adaptive filters, $w_1(n)$ and $w_2(n)$, respectively. The vectors of $C_1(n)$ and $C_2(n)$ are the adaptation gain, which are defined as

$$C_1(n) = \gamma_1(n)K_1(n) \quad (72.7)$$

$$C_2(n) = \gamma_2(n)K_2(n) \quad (72.8)$$

Where $\gamma_1(n)$, $\gamma_2(n)$ are likelihood variables, and $K_1(n)$, $K_2(n)$ are the dual Kalman gain, respectively. The likelihood variables and dual Kalman gain and the forward prediction errors variances are calculated using [10].

72.3 Analysis of Simulation Results

In this section, the simulation results of the various adaptive algorithms are discussed. The source signal is a speech signal taken from Noizeus database, wherein the sampling frequency is $f_s = 8$ kHz and the noise signal is taken from the same database that contains various noises like car, street, train noises at different SNR. The mixing signals $m_1(n)$ and $m_2(n)$ are generated by two cross-coupling impulse responses $h_1(n)$ and $h_2(n)$. The impulse responses, as shown in Fig. 72.1, are generated according to [3], by generating the random sequences according to equation, that is, $h(n) = Ae^{-Bn}$, where A is a scale factor that is linked to the variance of random sequence; in our sequence, in our case, the scalar A is taken to be equal to 1; and B is a damping factor, and is therefore linked to the reverberation time tr (i.e., $B = 3 \log(10)/tr$). The superiority of the speech signal is evaluated in terms of objectives measures like the segmental signal-to-noise ratio, the segmental mean square error. Figure 72.2 shows the original speech signal and its spectrogram of the original speech signal. Figure 72.3 shows the noisy speech signal and its spectrogram of the noisy speech signal. Figure 72.4 shows the enhanced speech signal and its spectrogram of the enhanced speech signal.

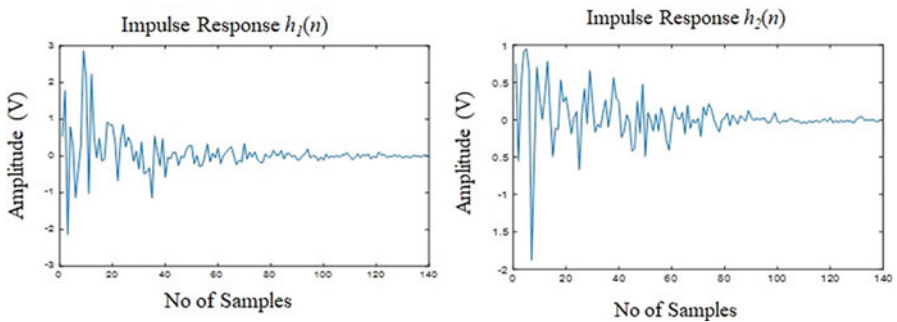


Fig. 72.1 Example of the simulated impulse response in left $h_1(n)$ and in right $h_2(n)$

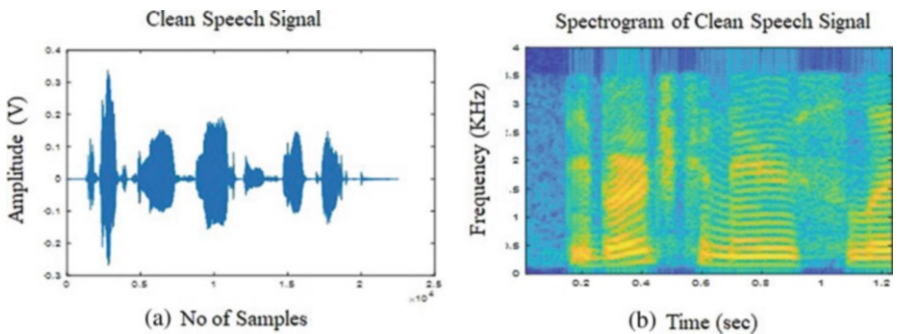


Fig. 72.2 (a) Original speech signal and (b) Spectrogram of the original speech signal

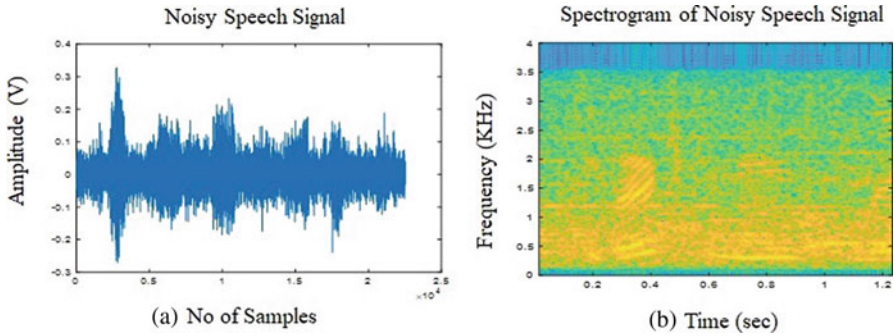


Fig. 72.3 (a) Noisy speech signal and (b) Spectrogram of the noisy speech signal

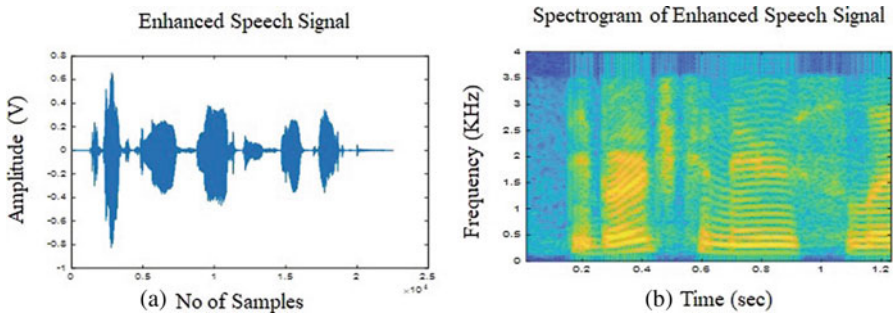


Fig. 72.4 (a) Enhanced speech signal and (b) Spectrogram of the enhanced speech signal

72.3.1 Segmental SNR (SegSNR) Evaluation

The SegSNR estimation is based on the following relation:

$$\text{segSNR}_{\text{db}} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left(\frac{\sum_{n=0}^{N-1} |s(n)|^2}{\sum_{n=0}^{N-1} |s(n) - u_1|^2} \right) \quad (72.9)$$

Where, $s(n)$ and $u_1(n)$ are the original and the enhanced speech signals. The variable M represents the number of segments and N represents the segment length. From the Figs. 72.5 and 72.6, we can observe that good behavior of DFNLMS in comparison with the LMS and NLMS algorithms.

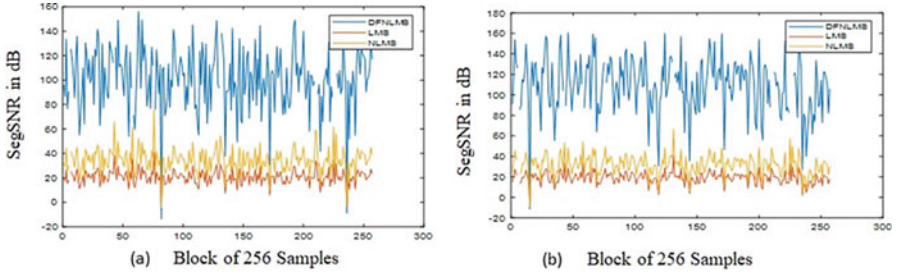


Fig. 72.5 SegSNR of the proposed algorithm and LMS and NLMS for the adaptive filter. (a) $L = 64$ at 0 dB; (b) $L = 64$ at 5 dB

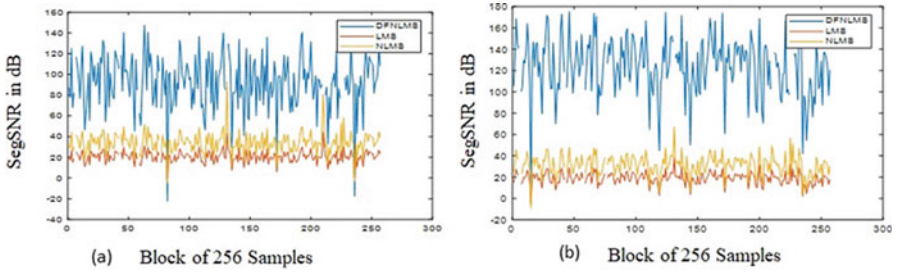


Fig. 72.6 SegSNR of the proposed algorithm and LMS and NLMS for the adaptive filter. (a) $L = 128$ at 0 dB (b) $L = 128$ at 5 dB

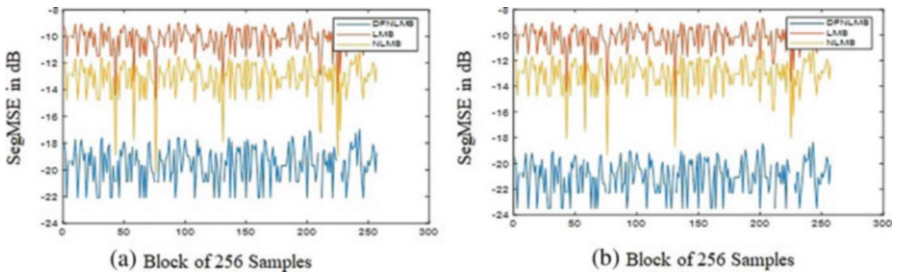


Fig. 72.7 SegMSE of the proposed algorithm and LMS and NLMS for the adaptive filter. (a) $L = 64$ at 0 dB (b) $L = 64$ at 5 dB

72.3.2 Segmental MSE (SegMSE) Evaluation

The segmental mean square error is given by the following relation:

$$\text{SegMSE}_{dB} = \frac{10}{M} \sum_{m=0}^{M-1} \log_{10} \left(\frac{1}{N} \sum_{n=0}^{N-1} |s(n) - u_1(n)|^2 \right) \quad (72.10)$$

where N is the time averaging frame length of the original and enhanced signals, and M is the number of segments. From the Figs. 72.7 and 72.8, we see the faster

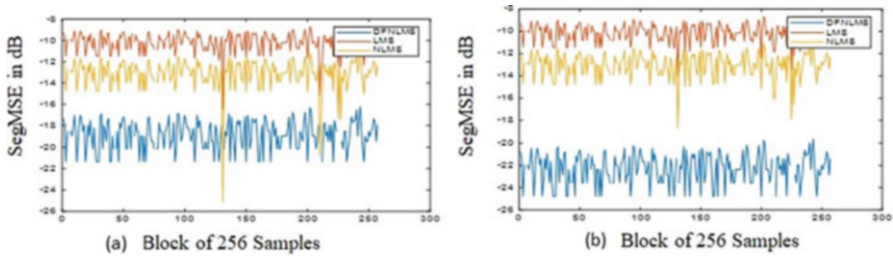


Fig. 72.8 SegMSE of the proposed algorithm and LMS and NLMS for the adaptive filter. (a) $L = 128$ at 0 dB (b) $L = 128$ at 5 dB

Table 72.1 SNR performance measure at 0 dB

S.No.	Noise at 0 dB	LMS	NLMS	DFNLMS
1	Car	15.95	26.65	41.10
2	Street	13.57	14.17	47.38
3	Airport	11.53	11.62	57.17

Table 72.2 SNR performance measure at 5 dB

S.No.	Noise at 5 dB	LMS	NLMS	DFNLMS
1	Car	32.94	38.19	52.34
2	Street	21.61	32.61	49.60
3	Airport	27.21	33.63	60.00

Table 72.3 SNR performance measure at 10 dB

S.No.	Noise at 10 dB	LMS	NLMS	DFNLMS
1	Car	37.23	52.50	74.32
2	Street	27.44	38.41	50.16
3	Airport	24.84	45.06	66.41

convergence speed performance of the DFNLMS with the classical LMS and NLMS, respectively.

To perform comparison evaluation, the speech enhancement algorithm is subjected to all noise types, and the performance is measured in terms of output SNR and it is given in Tables 72.1, 72.2 and 72.3.

The output SNR is given by,

$$\text{OutputSNR} = 10 \log_{10} \frac{\sum_{r=1}^N S^2(r)}{\sum_{r=0}^N [\hat{s}(r) - s(r)]^2} \text{ dB} \tag{72.11}$$

Where N is the number of samples; $s(r)$ is the clean speech signal; $n(r)$ is the noise; and $\hat{s}(r)$ is the enhanced speech signal.

Tables 72.1, 72.2, and 72.3 show the performance of signal-to-noise ratio at different noises such as 0, 5, and 10 dB. Comparing LMS and NLMS, DFNLMS performs better in all noise levels such as 0, 5, and 10 dB and achieves good SNR.

72.4 Conclusion

The comparisons of different adaptive algorithms like LMS, NLMS, and DFNLMs algorithms are discussed. Among the different types of adaptive filter, DFNLMs algorithm has shown a faster convergence speed with different noisy observations. The obtained result of DFNLMs is numerically stable and does not need any initial condition to converge toward the optimal solution to enhance the noisy speech signal. From the obtained results, we can say that the DFNLMs algorithm has shown the best performance in terms of the objective criteria like segmental SNR and segmental MSE and signal-to-noise ratio.

References

1. Thenua RK, Agarwal SK (2010) Simulation and performance analysis of adaptive filter in noise cancellation. *Int J Eng Sci Technol* 2(9):4373–4378
2. Thenua RK (2011) Hardware implementation of adaptive algorithms for noise cancellation. *International conference on network communication and computer*
3. Loizou PC (2013) *Speech enhancement: theory and practice*, 2nd edn. CRC Press, Boca Raton
4. Djendi M, Gilloire A, Scalart P (2006) Noise cancellation using two closely spaced microphones: experimental study with a specific model and two adaptive algorithms. In *IEEE International Conference ICASSP*, Toulouse, France, 3, pp 744–748
5. Rakesh P, Kumar TK (2015) A novel RLS adaptive filtering method for speech enhancement. *Electr Comput Energetic* 9(2):225–229
6. Djendi M, Henni R, Sayoud A (2016) A new dual forward BSS based RLS algorithm for speech enhancement. In: *International conference on engineering and MIS, ICEMIS 2016*, Agadir, Morocco
7. Djendi M, Bendoumia R (2016) Improved subband-forward algorithm for acoustic noise reduction and speech quality enhancement. *Appl Soft Comput* 42:132–143
8. Djendi M, Scalart P (2012) Double Pseudo affine projection algorithm for speech enhancement and acoustic noise reduction. In: *Proceeding of the IEEE EUSIPCO*, Romania, Bucharest, 1, pp 2080–2084
9. Ghribi K, Djendi M, Berkani D (2016) A new wavelet-based forward BSS algorithm for acoustic noise reduction and speech quality enhancement. *Appl Acoust* 105:55–66
10. Sayoud A, Djendi M, Medahi S (2018) A dual fast NLMS adaptive filtering algorithm for blind speech quality enhancement. *Appl Acoust* 135:10

Chapter 73

Android Malware Detection



Shymala Gowri Selvaganapathy, G. Sudha Sadasivam, Hema Priya N, Rajeshwari N, Dharani M, and K. Karthik

Abstract Smartphones and mobile tablets are rapidly becoming essential in daily life. Android has been the most popular mobile operating system since 2012. However, owing to the open nature of Android, countless malwares are intermingled with a large number of benign apps in Android markets that seriously threaten Android security. The botnet is an example of using good technologies for bad intentions. A botnet is a collection of Internet-connected devices, each of which is running one or more bots. The Bot devices include PCs, Internet of Things, mobile devices, etc. Botnets can be used to perform Distributed Denial of Service (DDoS attack), steal data, send spam and allow the attacker access to the device and its connection. To ensure the security of mobile devices, malwares have to be resolved. Malware analysis can be carried out using techniques like static, dynamic, behavioural, hybrid and code analysis. In this chapter, several machine learning techniques and classifiers are used to categorize mobile botnet detection.

Keywords Android · Android security · Botnet · Static · Dynamic · Hybrid · Malware detection · Machine learning techniques

Abbreviation

DDoS	Distributed Denial of Service
CCTree	Categorical Clustering Tree
APK	Android Package
SVM	Support Vector Machine
ELM	Extreme Learning Machine
SLFN	Single-Layer Feedforward Neural Network
CNN	Convolutional Neural Network

S. G. Selvaganapathy (✉) · G. S. Sadasivam · Hema Priya N · Rajeshwari N · Dharani M · K. Karthik
PSG College of Technology, Coimbatore, Tamil Nadu, India

73.1 Introduction

In the present era of hypersonic technology proliferation, predominantly everyone is gazing with the handheld devices like smartphones, tablets, laptops and that depicts technology’s fast-growing pace and interpret the force behind it all. Amidst the various mobile-operating systems, Android stands first as per the statistics [1] in Fig. 73.1. Despite its popularity, there are several security issues in Android which are caused by malwares, software which is specifically designed to disrupt, damage or gain authorized access to a computer system like Virus, Worms, Botnets, etc. So to provide security to mobile devices, malware has to be detected. Mobile botnets are type of botnets that target mobile devices, attempting to gain complete access to the device. Android malware detection and categorization can be done by several machine learning techniques like Support Vector Machine, Extreme Learning Algorithm, Logistics Regression and Convolutional Neural network.

73.2 Literature Survey

Given the potential fallout of installing malware on a personal computer or on a mobile device, there have been many proposed solutions for malware classification. Recently, researchers have been attempting to use machine learning and deep learning models to improve malware classification.

Ahmad Karim et al. [2] has proposed a framework which learns to distinguish applications having C&C functionality from malicious corpus through dynamic analysis of android applications. The framework were purely based on machine learning techniques that can classify applications based on various features collected

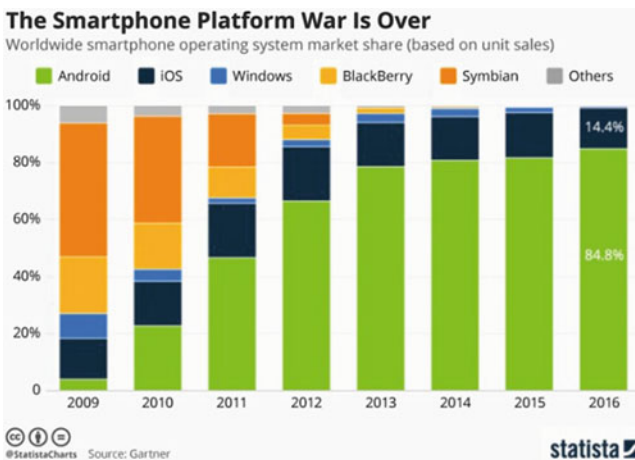


Fig. 73.1 Android statistics

at runtime through dynamic analysis of datasets. They proposed a learning technique by inducing Artificial Neural Networks back-propagation method.

AndiFitriah Abdul Kadir et al. [3] proposed a system of a deep analysis of Command and Control (C&C) and built-in URLs of Android botnets through static and dynamic analysis of the dataset Anubis consisting of 14 families. The dynamic analysis was conducted using Anubis, a web-based malware analysis tool. During the analysis, hidden encryption keys stored within the samples that allow revealing previously unknown features of Android botnets are identified, which are currently used to avoid detection.

DREBIN [4] a lightweight method for detection of Android malware that infers detection patterns automatically and enables identifying malware directly on the smartphone. DREBIN statically inspects a given Android application and extracts different feature sets from the application's manifest and dex code using a dataset of real Android applications. Another limitation which follows the use of machine learning has the possibility of mimicry and poisoning attacks.

Wenyi Huang et al. [5] has modelled a new deep learning malware classification architecture, MtNet, which shows for the first time that deep learning a modest improvement compared to a shallow neural architecture. The MtNet architecture also employs rectified linear unit (ReLU) activation functions and dropout for the hidden layers. However, MtNet has also vulnerable to the recently reported attack for deep neural networks proposed by Papernot, et al. [6].

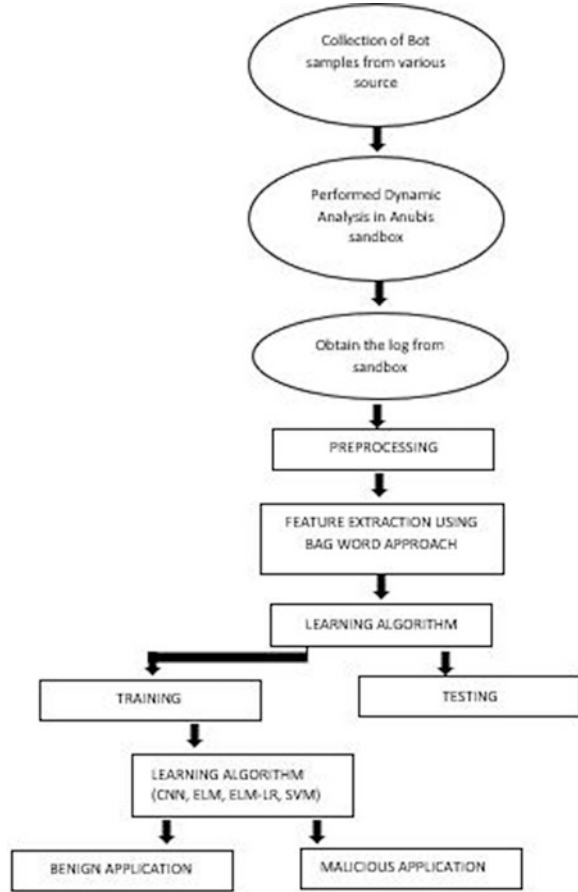
Antonio La Marra et al. [7] has proposed MalProfiler, a novel framework based on the Categorical Clustering Tree (CCTree) algorithm, to perform both automated grouping and classification of malicious apps in behavioural classes. The proposed framework clusters large amount of unlabelled malicious applications into smaller homogeneous sets. The analysis is based on a set of static features, extracted directly from the app's APK files, which are representative of both the app structure and of the performed behaviours.

Another interesting behavioural detection framework is proposed in Bose et al. [8] to detect mobile worms, viruses and trojans. The method implies a temporal logic approach to detect malicious activity in time. The ability of this framework to detect new types of malware is still uncertain as it requires a process of specifying temporal patterns for the malicious activities.

73.3 Proposed Work

The proposed system utilizes android botnet samples, perform dynamic analysis extended by machine learning techniques on the application files and categorize them into respective botnet families (Fig. 73.2).

Fig. 73.2 System design of proposed model



73.3.1 Dataset Collection

An integrated dataset consisting of early and mature versions of Android botnets is generated by collecting samples representing 14 botnet families. The bot samples are collected from Android Genome Malware project [11], malware security blog [12], VirusTotal [13] and samples provided by a well-known anti-malware vendor. This dataset includes 1929 samples of Android package (APK) spanning a period from 2010 to 2014. It also covers a large number of existing Android botnets, which reflect the status of Android malware. The bot samples from different sources representing 14 families are stated in Table 73.1 [9] and the various commands and controls (C&C) they use. Dynamic analysis on bot samples was performed using the dynamic analysis toolkit *Anubis* [10]. Results from the dynamic analysis tool are analysed for further evaluations.

Table 73.1 Botnet families

Botnet family	C&C type	Propagation and attack types										Trojanized Application	
		Backdoor	Data theft	Drive by down load	Exploit technique	Infected SMS	Mobile banking attack	Ransom ware	Repackaged application	Social Engineer			
Wroba	SMS/ HTTP					YES	YES						YES
Pletor	SMS/ HTTP							YES					YES
Sandroid	SMS									YES			YES
Not compatible	HTTP			YES	YES								
Miso SMS	Email		YES		YES								YES
Bmaster	HTTP		YES		YES						YES		
Root smart	HTTP		YES		YES						YES		
Tiger bot	SMS	YES	YES										YES
Ansver bot	HTTP		YES			YES						YES	
Droid dream	HTTP		YES	YES	YES						YES		YES
Nicky spy	SMS		YES									YES	
Pjapps	HTTP											YES	YES
Geinimi	HTTP		YES	YES								YES	
Zitmo	SMS					YES				YES		YES	

73.3.2 Methodologies

Support Vector Machines Support vector machines, because of their inherent classification capability, have been widely used in classification purposes. The major learning phases in SVM include

- (i) Mapping the features onto a large vector space.
- (ii) Using the standard methods to determine the separation between the class labels.

SVMs are inherently two-class classifiers. The traditional way to do multiclass classification with SVMs is to build One Vs Rest classifier approach.

OneVsRest classifier creates multiple binary classification models, merges the models and then optimizes the algorithm for each class. One advantage of this approach is its interpretability. Since each class is represented by one classifier only, it is possible to gain knowledge about the classes by analysing its corresponding classifier. This is the most commonly used algorithm for multiclass classification.

Extreme Learning Machines Extreme Learning Machines (ELM) are single hidden-Layer Feedforward Neural Networks (SLFN), where only the output weights are optimized, and all the weights between the input and hidden layer are assigned randomly.

SLFN with N hidden layers and with arbitrarily chosen input weights can learn N distinct observations with arbitrarily small error. Unlike the most practical implementations that all the parameters of the feedforward networks need to be tuned, we may not necessarily adjust the input weights and first hidden layer biases in applications. Some simulations have shown that this method not only makes learning extremely fast but also produces good generalization performance.

ELM learning speed can be thousands of times faster than traditional feedforward network learning algorithms like back-propagation algorithm while obtaining better generalization performance. Different from traditional learning algorithms to the proposed learning algorithm not only tends to reach the smallest training error but also the smallest norm of weights.

ELM Algorithm

Input: A training dataset

Activation function (\cdot), and the number of hidden nodes M;

Output: A weights matrix

1. Randomly assign input weights W_j and biases b_j , $j = 1, 2, \dots, M$
2. Calculate the hidden layer output matrix H
3. Calculate output weights matrix $\hat{\beta} = H \cdot T$

Step1: Let $S = 0$.

Step2: Partition T into training set T_r and testing set T_e , i.e. $T = T_r \cup T_e$

- Step3: Train a probabilistic SLFN with ELM on Tr.
- Step4: While testing accuracy less than a predefined value, do
- Step5: For $\forall x \in Te$ calculate posterior probabilities $p(\omega_k|x)$ with trained SLFN;
- Step6: For $\forall a_i \in A$ let the weights $W_{ji} = 0$ ($J = 1, 2, M$) retrain new probabilistic SLFN with ELM on Tr;
- Step7: probabilities SLFN; For $\forall x \in Te$ recalculate posterior $p(\omega_k|x)$ with trained new probabilistic.
- Step8: Calculate $R(a_i) = \sum_{k=1}^c \sum_{x \in Te} | p(\omega_k|x) - p'(\omega_k|x) |$;
- Step9: Calculate $a_* = \text{argmax}\{R(a_i)\}$;
- Step10: Let $S = S \cup \{a_*\}$

ELM with its higher scalability and less computational complexity not only unifies different popular learning algorithms but also provides a unified solution to different practical applications like regression, binary and multiclass classifications. The activation function like sine, Gaussian, sigmoidal, etc., can be chosen for hidden neuron layer and linear activation functions for the output neurons.

Logistic Regression Logistic regression is used for a different class of classification problems. Classification is all about portioning the data into groups based on certain features. In regression models, hypothesis function can be written as $Y = C^T X$, where

$$x = \begin{bmatrix} 1 \\ x^1 \\ x^2 \\ \dots \\ x^n \end{bmatrix} \text{ and } c = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix} \text{ where } X_i \text{ is the vector containing } i\text{th feature value for all}$$

the entries in dataset.

A sigmoid function is applied over the general known hypothesis function to get it into a range of (0, 1). The new activation function becomes like $Sg(y) = sg(C^T X) = \frac{1}{1 + e^{-C^T X}}$. Logistics regression is generally used where the dependent variable is Binary or Dichotomous and it measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities.

Convolutional Neural Network CNN inherits the traits of an ordinary neural network. CNN stacks the neurons in a three-dimensional fashion. To implement for Anubis dataset, the log files are embedded onto vector space using a pertained word to vector model trained on a large corpus of text data. Then, the vector embedding's are given to the convolution layer, which performs convolution using multiple filter sizes. The next layer is the pooling layer, which max-pools the result of the convolution layer into a set of feature vectors and adds dropout regularization. The final layer is the classification layer which uses a softmax activation function to perform the classification task among multiple layers.

73.4 Result Analysis

The above four methodologies are used to evaluate the performance on the application. The multiclass SVM was trained for 70% as training set; the algorithm was predicted for 30% as testing set; and the accuracy score was obtained. In order to overcome the drawbacks of multiclass SVM in optimizing the various parameters to obtain a better accuracy, SLFN algorithm with only three-layered architecture is employed where the accuracy is much reduced. So ELM with logistic regression is applied then the accuracy rate on classification is increased. Finally, the dataset is tested with CNN algorithm and accuracy is calculated and shown in Table 73.2. The flow graph visually represents the performance of the elm algorithm pipelined with the logistic regression approach. The output of SVM and ELM are represented as graph visually (Fig. 73.3, 73.4, 73.5 and 73.6).

Table 73.2 Comparison of accuracy scores

Algorithm	Activation function used	Accuracy score
Multiclass SVM	Linear SVC	90.1%
ELM with SLFN	Hyperbolic tangent function	36.5%
ELM with logistic regression	Sigmoid	93.2%
CNN	Softmax	96.52%

Fig. 73.3 Comparison of performance of SVM and ELM algorithms

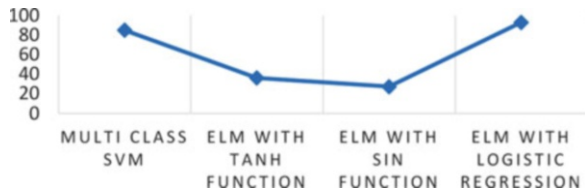
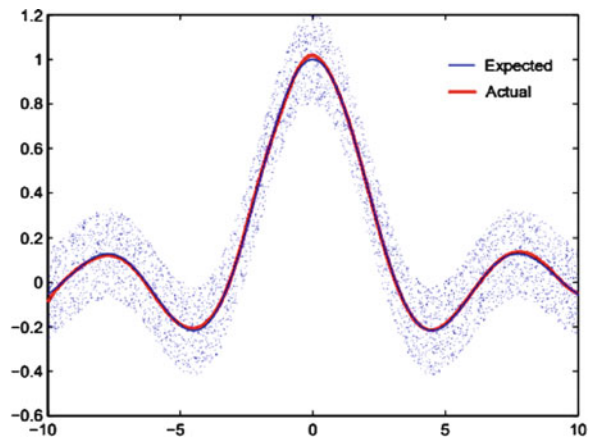


Fig. 73.4 Accuracy for ELM



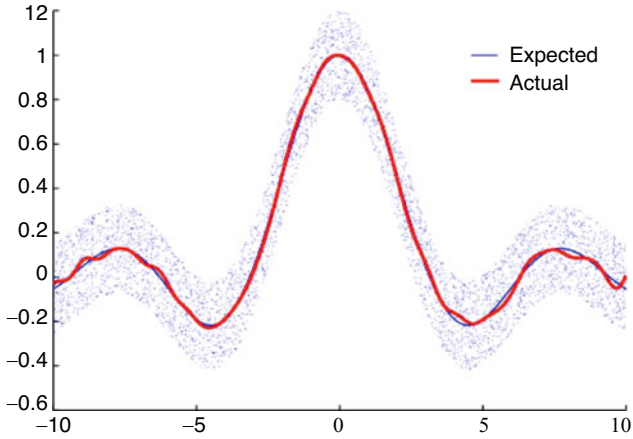


Fig. 73.5 Accuracy for SVM

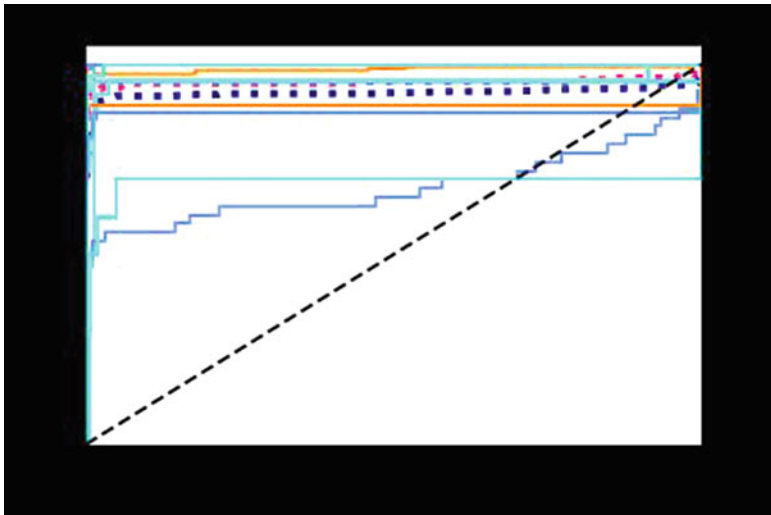


Fig. 73.6 ROC Curve of SVM

73.5 Conclusion

A framework is developed to analyse and detect Android-based mobile botnet application through dynamic analysis, augmented by machine learning technology. Extreme Learning Machines (ELM) are used for representation learning, and logistic

regression is used for multiclass classification. SVM and CNN are ML algorithms used to evaluate the proposed methodology. The future work can be to develop a hybrid on-device analysis for better categorization. For this purpose, we need to construct and implement the own sandbox with rich UI capabilities providing deep code coverage which can ultimately avoid all the deficiencies inherited from traditional dynamic analysis sandboxes.

References

1. <https://www.statista.com/chart/4112/smart-phone-platform-market-share/>
2. Papernot N, McDaniel P, Jha S, Fredrikson M, Celik ZB, Swamix A (2016) The limitations of deep learning in adversarial systems. In: IEEE European symposium on security and privacy
3. Kadir AFA, Stakhanova N, Ghorbani AA (2015) Android botnets: what URLs are telling us. in Springer
4. Bose A, Hu X, Shin KG, Park T (2008) Behavioral detection of malware on mobile handsets
5. MtNet: A Multi-Task Neural Network for Dynamic Malware Classification
6. MalProfiler: Automatic and Effective Classification of Android Malicious Apps in Behavioral Classes
7. <http://scikit-learn.org/one-vs-rest-classifier>
8. <http://medium.com/towards-data-science/activation-functions-and-its-types-which-is-better>
9. <http://www.unb.ca/cic/datasets/android-botnet.html>
10. Karim A, Salleh R, Shah SAA (2015) DeDroid: a mobile botnet detection approach based on static analysis. In IEEE
11. Liu L (2008) BotTracer: execution-based bot-like malware detection. In: ISC '08 proceedings of the 11th international conference on information security
12. Divita J, Hallman RA (2017) An approach to botnet malware detection using Nonparametric Bayesian Methods. In ACM

Chapter 74

Test Data Compression Methods: A Review



S. Rooban and R. Manimegalai

Abstract Integrated circuit (IC) applications have become viable, reliable, and cheaper with deep-submicron technologies in VLSI industry. SoC (system on chip) is a microchip, which holds the necessary hardware and software to implement various functions onto a single chip. An IC should be tested during design process to check the correctness of the design and also to check the functionality of the design after fabrication. Testing cost shares almost half of the manufacturing expenditure. Achieving high-test quality in reduced VLSI geometry increases the complexity of the testing methods because of huge volume of test data. Automatic testing equipment (ATE) reduces testing efforts, but it has limited memory in comparison with the huge volume of the test data. One of the methods to reduce the burden of ATE is test data compression. The advantages of test data compression are as follows: (i) reduction in memory requirement for ATE and (ii) reduction in testing time. Though many advanced algorithms are used for testing of VLSI circuits, most of them are expensive in terms of test data volume and power. This survey discusses all the test data compression methods, and it helps the researchers in testing domain.

Keywords VLSI testing · Test data compression · Decompressors

Abbreviations

ATE Automatic testing equipment
CSR Cyclic shift register
ATPG Automatic test pattern generator

S. Rooban
KLEF, Guntur, Andhra Pradesh, India
e-mail: rooban@kluniversity.in

R. Manimegalai (✉)
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India
e-mail: drmm@psgitech.ac.in

DFT	Design for testability
EFDR	Extended frequency-directed run-length
FDR	Frequency-directed run-length
FSM	Finite-state machine
IC	Integrated circuits
LFSR	Linear feedback shift registers
RBR	Run-based reordering
SoC	System on chip
VIHC	Variable-length input Huffman coding

74.1 Introduction

With increasing popularity of SoC circuits, the cost for testing is increased and becomes a serious challenge. Inaccessibility of embedded cores during the time of testing, test development, and application of huge volume of test data increases the testing cost. Implementation of several new nanomaterials in chip fabrication reduces the cost for single transistor but leads to several manufacturing defects. Rise in volumes of test data increases the cost of testing, energy usage, and test storage. Several industrial techniques are developed by experts to enable a compatible ratio between tests and production cost while ensuring acceptable levels of test quality. Test time and size of test data are reduced by the utilization of chip design with test data compression capabilities and to achieve effective usage of tester channels. Effective design necessitates the power aware aspect for improving the cost-effectiveness by compressing the test data to significant levels for system-on-chip testing. This paper discusses different designs, algorithms, and methodologies that are available in the literature for reduction of test data and power.

74.2 Test Data Compression Strategies

The past decade observed compression of test data as an effective method for reducing the volume of test data and test application time. In order to work with the increasing clock frequencies in a contemporary design, a low fill rate of care bits is exploited due to the result of combinational logic [1]. Hence, the fill rate of the every determined test pattern did not exceed 3–5% of total number of bits [2]. Test stimuli for compressed test data are stored in the tester memory and are transferred into scan chains through the on-chip decompressor. The three broad classifications of test data compression strategies in view of test pattern encoding procedures are as follows: (i) linear decompression-based schemes, (ii) code-based test data compression schemes, and (iii) broadcast-scan-based schemes [3].

74.3 Linear Decompression-Based Schemes

Decompressor with XOR gates and flip-flops is called as linear decompressor [4]. ATE produces a compressed test vector of q bits which is lesser than the input test vector of p bits. This q bits get the original p bits after passing through the decompressor by a Boolean matrix $A_{p \times q}$ where A is a characteristic matrix [5]. Linear equation is formed by characteristic equation, test vector, and free variable. Linear decompressor encoding capacity depends on the number of free variables in the linear equation. The optimality of linear decompressor is measured by how close its encoding efficiency is to 1.

74.3.1 *Combinational Decompressors*

The first method in linear test data compression series is combinational decompressors, which primarily depend upon linear operations and use linear XOR logic [6]. For example, a four input combinational decompressor can drive eight scan chains. When the total number of scan chains is more than the input channels, each channel feeds more than one scan chain. Because of the absence of memory element, the input is passed without any delay. For the uniformly specified bits, combinational decompressor gives acceptable results. In any case, test pattern that contain don't care bits can be grouped together. Highly specified scan slices are handled by adding additional input channels [7].

74.3.2 *Sequential Decompressor*

Sequential linear decompressors are connected sequentially to each core and are formed by linear finite-state machines (FSM) [8]. The various state machines are ring generator, Linear Feedback Shift Registers (LFSRs). Sequential linear decompressors having the advantage of increased encoding efficiency are compared to combinational decompressors. Encoding of sequential linear decompressor is attained by the free variable in previous clock and is used for the free variable in present clock [1]. This compressor does not require a separate CSR (cyclic shift register) and it reduces the hardware overhead. A perfectly designed sequential decompressor produces an encoding efficiency, which is greater than 0.95. FSM is connected between the XOR network and the decompressor to encode the test vector at the present clock cycle [9, 10]. Sequential decompressor is further divided into fixed-length sequential linear decompressor and variable-length sequential decompressor. Fixed length is also called as static reseeding [11]. Size of the LFSR is greater than the specified bits in the test cube. LFSR seed is created to encode the test cube. Variable-length sequential linear decompressor has the

capability to improve the encoding efficiency by increasing the free variables. It uses the maximum possible free variables for encoding [12].

74.3.3 Combined Linear and Nonlinear Decompressor

The possibility of occurrences of all the limitations in linear decompressor and the nonlinear decompressor are avoided by combining both the linear and nonlinear decompressor techniques [13]. The limitation of reduction in bit storage is caused by linear decompressor using statistical code. Performing the linear decompression with dictionary coding generates the decoding by linear decompressor when the code is missed in dictionary [14]. The similar nonlinear scan chains are incorporated into the linear decompressor [15]. Determining the association between the scan chains is complex for more scan length and huge volume of test patterns.

74.4 Code-Based Test Data Compression Techniques

Many previously proposed test compression techniques make use of familiar data compression algorithms developed earlier in the framework of information theory. These sets of methods do not consider the statistical recurrences of bits in the given test data. The majority of popular methods are based on code-based methodologies that replace repeated bits of test data with the help of code words. An on-chip decompression procedure rebuilds the original set of test data from the code words by reversing it into sets of symbols corresponding to it [16].

74.4.1 Dictionary-Based Code

The basic idea of dictionary-based approach is that utilizing the advantage of commonly occurring bit sequences by using a dictionary. The original test patterns are divided into n -bit blocks. These blocks are called symbols. Dictionary-based scheme uses fixed-length code words to represent each symbol. The frequently occurred bit sequences are swapped with code word in the dictionary. The basic complete dictionary-based compression scheme is described in [17]. To form a complete dictionary, the size of the dictionary is increased that causes complexity in decompressor design. Size of the dictionary is allocated in partial dictionary coding [18] by the user and depends on the availability of the area in decompressor, and it overcomes the drawback of size constraint in complete dictionary coding. Improvement in dictionary-based techniques is attained by considering mismatches. It finds out each bit in the patterns that are different in minimum bit positions in comparison with other patterns. Those differences are transferred to the compressed

bits and are added in the dictionary. Correction network for flipping the bit is added with partial dictionary to make the dictionary entry into the required scan slice [17]. Reduction in dictionary size is achieved by correction network [17]. A technique LZ77 that reduces 71% power consumption without affecting testing time and compression efficiency is proposed [14]. Combination of dictionary coding and partial LFSR reseeding is used to achieve increased compression ratio. It also uses matrix calculation for reducing testing time.

74.4.2 *Huffman Code*

It uses variable-length code words to represent the partitioned blocks of fixed size in original test patterns. It is also called statistical coding. The most frequently occurred symbols are replaced with shortest code word, and less frequently occurred symbols are swapped with higher length of code word. Thus it reduces the code word length average. Huffman code [19] is obtained by forming Huffman tree in which the value of binary bits flow from root node to leaf node produces the code word.

Selective Huffman coding [20] reduces the complexity and size of the decoder, only the most frequently appeared symbols are considered, and the remaining symbols are used as it is without any compression. By increasing the symbol size of considerable symbols, high compression ratio is achieved even not considering less frequently appeared symbols for compression, because full Huffman code requires the decoder with $2^n - 1$ states to decode all the encoded n -bit symbols results in increased silicon area for decoder. To identify the selected codes for encoding, an additional bit is added. The bit value of “1” in that added bit represents it is considered for encoding in the presence of “0” represent which is not considered. This additional bit increases the overhead. The overhead caused by that extra bit is removed by adding an additional bit only to the un-encoded symbols. Thus reduction in one bit for most frequently appeared symbols produces better compression than the previous selective Huffman code.

74.4.3 *Run-Length Code*

It is a variable-to-fixed coding. The original test data is divided into variable-length symbols that depend on the runs of successive 1's or 0's and coded to fixed number of bits. For example, an original data of 10 bits 01001 10,001 can be encoded by using 2-bit run-length for runs of 0's (1 → 00, 01 → 01, 001 → 10, 0001 → 11) as 01 10 00 11 (original data (01001 10,001) 10 bits, and run-length coded data (01 10 00 11), 8 bits). In cyclic scan architecture, the run-length coding is implemented. Only the difference of bits in the current vector with previous test vector is applied. All the test vectors are ordered in such a way that they have minimum differences between subsequent vectors, and the required bits applied are minimum and reduce

the memory requirement. To increase the efficiency of compression ratio, the variable to fixed coding is modified to variable-to-variable coding.

The technique frequency-directed run-length (FDR) is used in [16] that utilizes variable-to-variable length coding. FDR codes are not disturbed by any discrepancy in the input data stream, and the drawback is that FDR code is based on encoding runs of 0's. To overcome the drawback of FDR, extended frequency-directed run-length (EFDR) is described in [21]. This EFDR code is based on encoding of both runs of 0's and 1's and produces higher test data compression. Variable-length input Huffman coding (VIHC) [22] is a combination of Huffman and FDR. Run-based reordering (RBR) [23] is based on EFDR codes. Scan power and area overhead are also reduced in RBR. Dividing the coded word into head section with fixed number of bits and tail section with variable number of bits produces highest compression ratio.

74.5 Broadcast-Scan-Based Schemes

The basic principle in this scheme [24] is to apply the same value to various scan chains. This broadcast-scan-based scheme is used in design for testability (DFT) and also in various test data compression architectures. Broadcast scan shares the pattern with independent circuits by connecting the inputs of the circuits to be shared in a 1–1 mapping method. This way of connection does not require any modification in the automatic test pattern generator (ATPG) program. Test compression is achieved by applying only one pattern to multiple circuits as if it were a single circuit. There is no reduction in fault coverage in broadcast scan when patterns are applied to independent circuits. This approach can also be applied to multiple scan chains of the same circuits if there are no dependencies between the sub-circuits driven by scan chains.

Drawbacks of applying broadcast scan to the multiple circuits that are dependent can be overcome by Illinois scan. In Shah and Patel [25], test patterns are applied with two modes, broadcast mode and serial scan mode. In broadcast mode, a scan chain is divided into multiple sub-scan chains. By the use of single shared scan, the same input pattern is shifted into all the sub-scan chains to identify the most faults in the circuits. Then the test patterns are applied in serial scan mode to identify the remaining faults for full fault coverage. One drawback of this scheme is that there is no compression ratio maintained, if only less faults are needed to be identified in serial scan mode.

Multiple channels are used to drive all the scan chain to reduce the number of patterns used by the serial scan mode. Increasing the channels makes shorter scan chain length and increases the number of faults detected by the broadcast scans. Identifying the suitable configurations, which detects all the faults with minimum number of channels, increases the efficiency of broadcast scan and improves compression ratio. A reconfigurable broadcast scan is used to reduce the number of channels used by the multiple input broadcast scan and to achieve high fault

coverage. There are two reconfigurations; they are static reconfiguration and dynamic reconfigurations.

In static reconfiguration [26], the fault to be identified is divided into various subsets. Each subset is identified with a single configuration. Once the first subset fault is identified, the configuration is changed to identify the next subset of fault. Static reconfigurations can only be applied to the new entry of test patterns. In [27, 29] the change in configuration is done during scanning a pattern dynamically, which provides more flexibility. It leads to better results with less number of channels. More control information is required in dynamic configuration for reconfiguring at the right time compared to static configurations. In [28], the constraint imposed on the decompressor is removed by adding the additional logic through virtual scan inputs. This scheme avoids fault coverage loss caused by the compression techniques. In virtual scan, the ATPG tool identifies the test patterns, which can make the decompressor to produce the effective dynamic compaction. By applying those test patterns, overall efficiency of testing scheme is improved.

74.6 Other Compression Techniques and Classification by DFT Vendors

The above three broad classifications of test data, in compression techniques, many researchers have proposed unique technique by combining two or more schemes. They are (i) LFSR-based techniques (ii) hybrid techniques, and (iii) predictive techniques. There are many companies Cadence, Mentor Graphics, Syntest, Synopsis, and Logic Vision which produce various compression tools. Those vendors classify the compression techniques into two groups based on the type of ATPG used. The two groups are linear-decompression-based schemes and broadcast-scan-based schemes. The various products developed by the vendors for linear-decompression-based schemes are ET Compression from Logic Vision, Test Kompress from Mentor Graphics, and SOCBIST from Synopsys. In this group test, compression is achieved by two steps. In the first step, normal ATPG is used to produce the test patterns and performs dynamic compaction and no aggressiveness in filling the test cubes increases the coverage. In second step, linear equations are used to map the external scan input ports to the internal scan chain inputs. The products for other groups are OPMISR + from Cadence, Virtual Scan and Ultra Scan from Syntest, and DFT MAX from Synopsys.

74.7 Conclusion

This paper presents a survey on test compression techniques in various schemes. Different power management techniques are proposed by researchers. The discussed methods addressed one or more of the following issues: encoding efficiency, test application time, tester storage requirement, and test power. Most of the techniques proposed for code-based test data compression scheme are easier to implement. Implementing these techniques with high compression efficiency is an interesting open problem.

References

1. Touba NA (2006) Survey of test vector compression techniques. *IEEE Des Test Comput* 23 (4):294–303
2. Rajski J (2008) Logic diagnosis, yield larning & quality of test. In: Proceedings of the IEEE international symposium on VLSI design, automation and test, Taiwan, pp 3–3
3. Varma P, Bathia S (1998) A structured test re-use methodology for core-based system chips. In: Proceedings of the international test conference, Washington, DC, pp 294–302
4. Balakrishnan KJ, Touba NA (2006) Improving linear test data compression. *IEEE Trans Very Large Scale Integr Syst* 14(11):1227–1237
5. Laung-Terng W, Cheng-Wen W, Xiaoqing W (2007) Review of VLSI test principles and architecture. *IEEE Des Test Comput* 24(2):198–199
6. Mitra S, Kim K (2006) XPAND: an efficient test stimulus compression technique. *IEEE Trans Comput* 55(2):163–173
7. Lee KJ, Chen JJ, Huang CH (1998) Using a single input to support multiple scan chains. In: Proceedings of the international conference on computer-aided design, San Jose, pp 74–78
8. Mrugalski G, Tyszer J, Rajski J (2004) Ring generators: new devices for embedded test applications. *IEEE Trans Comput Aided Des Integr Circuits Syst* 23(9):1306–1320
9. Krishna CV, Touba NA (2003) Adjustable width linear combinational scan vector decompression. In: Proceedings of the IEEE international conference on computer-aided design, San Jose, pp 863–866
10. Chandra A, Chakrabarty K (2003) Test data compression and test resource partitioning for system-on-a-Chip using frequency-directed run-length (FDR) codes. *IEEE Trans Comput* 52 (8):1076–1088
11. Krishna CV, Touba NA (2003) Hybrid BIST using an incrementally guided LFSR. In: Proceedings of the 18th IEEE symposium on defect and fault tolerance in VLSI systems, Boston, pp 217–224
12. Volkerink EH, Mitra S (2003) Efficient seed utilization for reseeding based compression. In: Proceedings of the VLSI test symposium, California, pp 232–237
13. Krishna CV, Touba NA (2002) Reducing test data volume using LFSR reseeding with seed compression. In: Proceedings of the international test conference, Baltimore, pp 321–330
14. Sun X, Kinney L (2004) Combining dictionary coding and LFSR reseeding for test data compression. In: Proceedings of the 41st IEEE design automation conference, San Diego, pp 944–947
15. Li L, Chakrabarty K, Kajihara S, Swaminathan S (2005) Efficient space/time compression to reduce test data volume and testing time for IP cores. In: Proceedings of the 18th international conference on VLSI design, Kolkatt, pp 53–58

16. Usha Mehta S, Kankar Dasupta S, Niranjana M (2010) Devashrayee: hamming distance based reordering and column-wise bit stuffing with difference vector: a better scheme for test data compression with run length based codes. In: Proceedings of the 23rd international conference on VLSI design, Bangalore, pp 33–38
17. Chandan G, Santanu C (2006) Power optimized dictionary coding for test data compression. In: Proceedings of the IEEE international conference on industrial technology, Mumbai, pp 2541–2545
18. Li L, Chakrabarty K, Touba NA (2003) Test data compression using dictionaries with selective entries and fixed-length indices. *ACM Trans Des Autom Electron Syst* 8(4):470–490
19. Huffman DA (1952) A method for the construction of minimum redundancy codes. *Proc Inst Radio Eng J* 40(9):1098–1101
20. Jas A, Dastidar JG, Mom-Eng J, Touba NA (2003) An efficient test vector compression scheme using selective Huffman coding. *IEEE Trans Comput Aided Des Integr Circuits Syst* 22(6):797–806
21. El-Maleh AH (2008) Efficient test compression technique based on block merging. *IET Comput Digit Tech* 2(5):327–335
22. Gonciari PT, Al-Hashimi B, Nicolici N (2003) Variable-length input Huffman coding for system-on-a-chip test. *IEEE Trans Comput Aided Des Integr Circuits Syst* 22(6):783–796
23. Wenfa Z (2007) Test data compression scheme based on variable-to-fixed-plus-variable-length coding. *J Syst Archit* 53(11):877–887
24. Lee K-J, Chen JJ, Huang CH (1999) Broadcasting test patterns to multiple circuits. *IEEE Trans Comput Aided Des* 18(12):1793–1802
25. Shah MA, Patel JH (2004) Enhancement of the Illinois scan architecture for use with multiple scan inputs. In: Proceedings of the annual symposium on VLSI, Lafayette, pp 167–172
26. Samaranyake S, Gizdarski E, Sitchinava N, Neuveux F, Kapur R, Williams TW (2003) A reconfigurable shared scan-in architecture. In: Proceedings of the 21st IEEE VLSI test symposium, Napa Valley, pp 9–14
27. Li L, Chakrabarty K (2004) Test set embedding for deterministic BIST using a reconfigurable interconnection network. *Trans Comput Aided Des Integr Circuits Syst* 23:1289–1305
28. Wang S, Gupta SK (2002) DS-LFSR: a BIST TPG for low switching activity. *Trans Comput Aided Des Integr Circuits Syst* 21(7):842–851
29. Rajsiki J, Tyszer J, Kassab M, Mukherjee N (2004) Embedded deterministic test. *IEEE Trans Comput Aided Des Integr Circuits Syst* 23(5):776–792

Chapter 75

Piripori: Morphological Analyser for Tamil



M. Suriyah, Aarthy Anandan, Anitha Narasimhan, and Madhan Karky

Abstract Morphology is the study of internal structure of word forms. A morphological analyser is a tool which gets the inflected, derived or compound forms of words as input and retrieves various grammatical implications of the input such as root word(s) and supporting morphemes. Prominent south Indian languages like Tamil, Telugu, Malayalam, Kannada and Tulu are agglutinative. In agglutinative languages, complex words are formed by joining morphemes together without changing the spelling or phonetics. With enormous amount of text data being generated daily in different platforms, it is important to have linguistic tools to analyse the same. Several approaches have been proposed for developing a Tamil morphological analyser. The approach adopted for Piripori Tamil morphological analyser is based on word-level morphological rules. This paper proposes a unique architecture for morphological analysis. The computational time taken by Piripori and the state-of-the-art morphological analyser have been analysed and compared with an input set of 137,144 words. It was observed that the proposed approach achieves faster computation when compared to the state-of-the-art analyser.

Keywords Piripori · Tamil morphological analyser · Rule-based morphological analyser · Natural language processing · Abbreviation · SVM Support Vector Machine

75.1 Introduction

Unicode has made internationalization and localization possible in a big scale. It supports almost 953 languages [1] and has enabled non-English content to be posted on the Internet. With the Internet becoming easily available these days, the volume of data generated through various platforms like social media, blogging and micro-

M. Suriyah · A. Anandan · A. Narasimhan · M. Karky (✉)
Karky Research Foundation, Chennai, Tamil Nadu, India
e-mail: inbox@karky.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_75

801

blogging sites, photo and video sharing applications, etc., has become huge. The data is thus generated with increased velocity, and veracity can be analysed for purposes like marketing, sentiment analysis, opinion mining, etc. Thus the Internet provides a warehouse of information which can be tapped by various agencies like companies, governments, etc., to aid better decision-making and policy framing. This paper proposes a morphological analyser for Tamil, which, apart from analysing the morphological implications of a given word, also translates it, thus aiding in better data handling and information retrieval.

Morphology involves the identification of morphemes, which are the smallest linguistic pieces with grammatical information [2]. For instance, the word “unbelievable” has three morphemes, “un”, “believe” and “able”. These three morphemes lose their meaning when divided further. Hence, morphology, in linguistics, is the study of the forms of words and the ways in which words are related to other words of the same language [3].

Morphological analyser is a component used to split a word to constituent morphemes and analyse the grammatical implication of each morpheme in the word. It is an indispensable component in natural language processing applications. In this paper, Piripori, a morphological analyser for Tamil, is studied. It is a rule-based system which works based on specific set of rules which handle nouns, verbs and compound nouns separately. It follows morpheme-based morphology.

75.2 Literature Review

Dhanalakshmi et al. propose a machine learning-based approach for Tamil morphological analysis which views the task as a classification problem [4].

The regular inflectional patterns in Tamil are modelled as a finite state automata for better accuracy in morpheme extraction by Sobha Lalitha Devi et al. [5]. A Tamil word is processed from right to left. When a suffix is accepted by a finite state automaton, the processing proceeds till it reaches the final state where the root word is found. If the morphemes are accepted at all states, then the word is taken as a valid word and its morphological analysis is returned.

To build a morphological analyser for classical Tamil, the input word is first checked to find out if it is a root word by itself. If the given word is not a root word, a rule-based approach is used to split the word to morphemes and obtain the analysis [6].

Mokanarangan et al. designed a morphological engine to generate all candidate analyses of a word using an annotated lexicon corpus and a repository of Tamil grammatical rules [7]. These candidates were input to an SVM classifier which used a high-frequency word list to predict the best analysis of the word. The features considered by this SVM were frequency-based scores, suffix, lexical labels and average length.

Jisha et al. attempt to build a Malayalam to Tamil machine translating system by using a Malayalam morphological analyser, a Malayalam Tamil bilingual dictionary and Tamil morphological generator [8].

Anand Kumar et al. [9] employ SVM for morphological analysis, thus making use of a machine learning-based approach. The point of concern with supervised learning methodologies is that the training data has to be voluminous and versatile to get reliable results. Additionally, when the rules governing morphological transformations are formed by analysing all the linguistic and grammatical features, grouped and streamlined effectively, the rule-based approach works faster with better precision compared to machine learning-based morphological analysers.

75.3 Foundation of Piripori

Piripori is a rule-based system which works based on basic inflectional principles of Tamil language at word level grouped in a three-level classification. The classes are patterns, rules and exceptions.

75.3.1 Patterns

These are the affixes added to the root form of a word to form different morphological variations of the word considered. These patterns contain morphological information of number and case for nouns; in the case of verbs, they denote tense, number, case, person and gender.

Table 75.1 demonstrates some patterns applied to the [மலர் | malar | flower]. Here, [மலர் | malar | flower] is the root word which is a noun in singular form.

75.3.2 Rules

When a root word combines with a specific pattern (suffix with some morphological information), the former may have a change in one or more letters, or may have one or more letters added to it to form the final inflected form. These transformations are governed by grammatical principles described in [12–14]. These are classified as “rules” under each pattern, i.e., each pattern has an associated set of rules governing the transformations which may occur when the pattern combines with a word to form the inflected form of the latter. For example, the [முள் | muL | thorn] has [ள் | L] as the last letter. When [முள் | muL | thorn] is pluralized with the suffixing of the pattern [கள் | kaL], the “ள்” has to be transformed to [் | t]. Some rules are tabulated in Table 75.2.

Table 75.1 Sample patterns

Pattern	Implication of the pattern	Root	Inflected form
[கள் kaL]	<plural>	[மலர் malar flower]	[மலர்கள் malargaL flower<plural>]
[களில் kaLil]	<plural> <locative case>	[மலர் malar flower]	[மலர்களில் malargaLil flower <plural> <locative case>]
[களிலிருந்து kaLilirundhu]	<plural> <ablative case>	[மலர் malar flower]	[மலர்களிலிருந்து malargaLilirundhu flower<plural> <locative case> <ablative case>]

Table 75.2 Sample rules

Pattern	Root noun	Rule	Inflected form	Inflectional split
[கள் kaL]	[முள் muL thorn]	[< > ள் t > L]	[முட்கள் mutkaL thorn<plural>]	[முள் + கள் muL + kaL]
	[சொல் chol word]	[< > ல் R > l]	[சொற்கள் choRkaL word<plural>]	[சொல் + கள் chol + kaL]

75.3.3 Exceptions

These are transformations which do not follow any regular structure with respect to the patterns and rules discussed earlier. Some exceptions are tabulated in Table 75.3.

The verb [வா | vaa | come] combines with the pattern [ந்தான் | ndhaan] to form [வந்தான் | vandhaan | he came], i.e. the alphabet [வா | vaa] changes to [வ | va]. The same verb combines with the pattern [கிறான் | kiRaan] to form [வருகிறான் | varugiRaan | he comes]. In this inflection, the alphabet [வா | vaa] changes to [வ | va] and the alphabet [ரு | ru] gets added as well.

It can be observed from the above discussion that the morphological rules governing word inflections can be consolidated in three levels, namely, patterns, rules and exceptions. Two hundred eighty noun patterns and 880 verb patterns were collected. Each pattern is given a specific ID. All the rules applicable for a specific pattern are grouped with the ID allotted for the pattern. The number of rules for nouns, verbs and compound words is 50, 124 and 54, respectively.

Table 75.3 Sample exceptions

Root Verb	Pattern	Implication of the pattern	Inflected form of the word	Intermediate inflectional split
[வா vaa come]	[ந்தான் ndhaan]	<third person> <singular> <past tense> <masculine>	[வந்தான் vandhaan he came]	[வ + ந்தான் va + ndhaan]
[வா vaa come]	[கிறான் kiRaan]	<third person> <singular> <present tense> <masculine>	[வருகிறான் varugiRaan he comes]	[வரு + கிறான் varu + kiRaan]

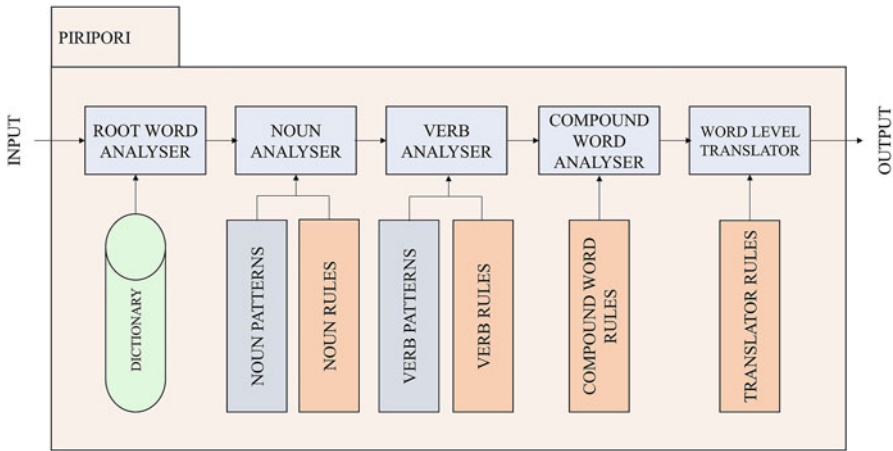


Fig. 75.1 Architecture of piripori

75.4 Architecture of Piripori

Piripori has five modules which make use of a dictionary of root words, the patterns and rules governing nouns and verbs, rules for compound word analysis and rules for translation at word level. The input to the system is any word – it may be a root word or a derived, inflected or compound word. The output is the morphological analysis of the input along with its word-level English translation. The architecture of the system is illustrated in Fig. 75.1.

- *Root word analyser*: When a word is given as input to the analyser, this module checks if the input is a root word by itself. It makes use of a dictionary of Tamil words for the above-mentioned purpose.
- *Noun analyser*: This module checks if the intermediate word segments obtained during analysis are valid noun suffixes, i.e., patterns [4]. If the segment matches with any of the noun patterns, the rules pertaining to that particular pattern are applied to the word and the morphological analysis of the word is obtained.
- *Verb analyser*: Just as the previous module, the verb analyser checks if the intermediate word segments are valid verb patterns. If there is a match, the rules associated with that specific verb pattern are applied to the word to obtain the analysis.
- *Compound noun analyser*: The compound noun analyser module applies rules for compound nouns on the given input word to obtain its morphological analysis. For example,
 - [வனமகன் | vanamagan | forest dweller] ← [வனம் | vanam | forest] + [மகன் | magan | son]
- *Word-level translator*: This module returns the translation of the input word by using the rules for translation.

The input word is checked if it is a root word itself. If yes, the word's part of speech and translation are returned. If not, the word is scanned from right to left. The scanned part is checked for it being a valid noun or verb pattern. If the scanned part is a valid pattern, the rules pertaining to it are applied to find the root word. If the entire scan is unsuccessful, the word is checked for compound word criteria. The number of nouns and verbs together is 138,212; these can be inflected in various forms by affixing different patterns in various combinations. Hence Piripori has a vast coverage over Tamil words.

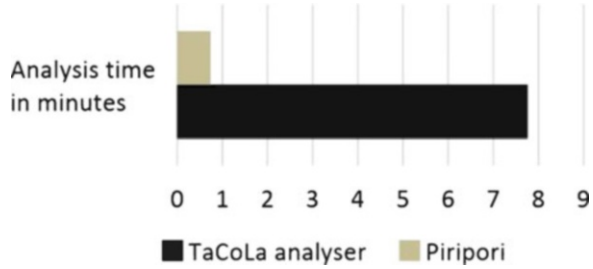
75.5 Results and Discussion

75.5.1 Analysis Time

A list of 137,144 Tamil words were taken from various sources in the internet and were used to compare the analysis times of the Piripori and TamilComputingLab morphological analysers [10]. Table 75.4 illustrates the results of the comparison. While TamilComputingLab analyser takes more than 4 min, the time taken by Piripori is just around one fourth of a minute. Fig. 75.2 illustrates the difference in the execution times.

Table 75.4 Comparison of analysis times

	Analysis time in nanoseconds	Analysis time in minutes
TaCoLa analyser	261,943,716,203	4.365728603
Piripori	14,597,149,487	0.243285825

Fig. 75.2 Comparison of analysis times**Table 75.5** Data for precision and recall calculation

	TaCola	Piripori
No. of nouns correctly classified	85	110
No. of verbs correctly classified	20	25
No. of nouns classified as verbs	9	0
No. of verbs classified as nouns	12	5

75.5.2 Precision and Recall

Precision is the fraction of retrieved words that have been correctly analysed. Recall is the fraction of words that have been correctly analysed.

To calculate precision and recall, a standard POS tagged document [11] was taken from LDCIL website and the nouns and verbs in the document were collected to two different files. There were 121 nouns and 43 verbs in the file, totalling to 164. The results of the same are tabulated in Table 75.5.

The following formulae are used for precision and recall calculation.

$$\text{Precision}_{\text{noun}} = (\text{No. of nouns correctly analysed}) / (\text{No. of words analysed as nouns})$$

$$\text{Precision}_{\text{verb}} = (\text{No. of verbs correctly analysed}) / (\text{No. of words analysed as verbs})$$

$$\text{Recall}_{\text{noun}} = (\text{No. of nouns correctly analysed}) / (\text{No. of nouns})$$

$$\text{Recall}_{\text{verb}} = (\text{No. of verbs correctly analysed}) / (\text{No. of verbs})$$

The precision and recall values of nouns and verbs are listed in Table 75.6. The graphical representation of the same is in Fig. 75.3.

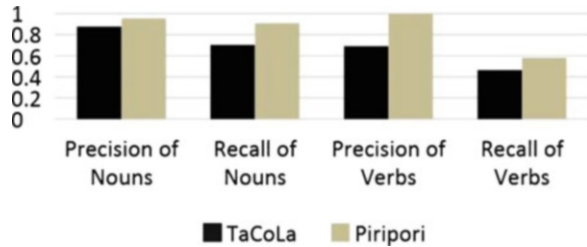
75.6 Conclusion

The grammatical rules of word inflection in Tamil discussed in [12–14] are consolidated in a three-level classification. The concepts listed are used to build a rule-based morphological analyser “Piripori”. It handles compound nouns in a better manner than other Tamil morphological analysers. The system also translates the

Table 75.6 Precision and recall of nouns and verbs

	TaCoLa	Piripori
Precision of nouns	0.87628866	0.9565217
Recall of nouns	0.702479339	0.9090909
Precision of verbs	0.689655172	1
Recall of verbs	0.465116279	0.5813953

Fig. 75.3 Comparison of precision and recall



word to English. The coverage of the analyser is vast due to the patterns and rules arrangement. It is also found to be having lesser analysis time than the TamilComputingLab morphological analyser [10]. If a word has multiple morphological analyses, all are listed as outputs.

In the case of a word having multiple analyses, some probabilistic or machine learning approach can be employed as in [7] to single out the best result. This system can be adapted to build a machine translation. Hence Piripori can be employed in various Tamil language applications for efficient and faster morphological analysis of words.

Acknowledgement The authors thank Subalalitha C.N., Kanniammal R., Manju Bashini P., Elanchezhian K., Rajapandian C., Swathi P.G. and Sindhuja G. for their suggestions and corrections.

References

1. The Unicode consortium [Online]. Available: https://unicode.org/cldr/charts/latest/supplemental/languages_and_scripts.html
2. University College Dublin [Online]. Available: <http://www.ucd.ie/artspgs/introoling/Aronoffmorphology.pdf>
3. Anderson SR (2003) Morphology. In: Encyclopedia of cognitive science. MacMillan, London
4. Dhanalakshmi V, Anandkumar M, Rekha RU, Arunkumar C, Soman KP, Rajendran S (2009) Morphological analyzer for agglutinative languages using machine learning approaches. In: Proceedings of international conference on advances in recent technologies in communication and computing, pp 433–435
5. Devi SL, Marimuthu K, Ram RVS, Bakiyavathi T, Amudha K (2013) Morpheme extraction in Tamil using finite state machines. Presented at forum for information retrieval evaluation, New Delhi, India, December 4–6, 2013

6. Akilan R, Naganathan ER (2015) Morphological analyser for classical tamil text: a rule based approach. *ARNP J Eng Appl Sci* 10(20):9325–9330
7. Mokanarangan TP, Megala U, Nilusija N, Dias G, Jayasena S, Ranathunga S (2016) Tamil morphological analyzer using support vector machines. In: *Proceedings of international conference on applications of natural language to information systems*, Salford, UK, pp 15–23
8. Jayan JP, Rajeev RR, Rajendran S (2015) Morphological analyser and morphological generator for Malayalam - Tamil machine translation. *Int J Comput Appl* 13(8):15–18
9. Anand Kumar M, Dhanalakshmi V, Soman KP, Rajendran S (2010) A sequence labeling approach to morphological analyzer for Tamil language. *Int J Comput Sci Eng* 2(6):1944–1951
10. Anandan P, Saravanan K, Parthasarathi R, Geetha TV (2002) Morphological analyzer for Tamil. In: *Proceedings of international conference on natural language processing*
11. Linguistic Data Consortium for Indian Languages [Online]. Available: <http://www.ldcil.org/standardsTextPOS.aspx>
12. Maanikkavasaasakan (2010) *Tholkaappiyam*. Uma Padhippaagam, Chennai
13. Manikkam A (1988) *Nannool Kaandigaiyurai*. Poompuhar Padhippagam, Chennai
14. Muhammad SN (2013) *Nalla Tamizh Ilakkanam*. Adayalam Padhippagam, Chennai

Chapter 76

A Comprehensive Survey on Strategies in Multicore Architectures, Design Considerations and Challenges



R. Radhika, N. Anusha, and R. Manimegalai

Abstract CMOS technology in contemporary period is enhanced with advanced features and compatible storage system. Advanced CMOS technology provides functional density, increased performance, reduced power, etc. System-on-chip (SoC) technology provides a path for continual improvement in performance, power, cost, and size at the system level in contrast with the conventional CMOS scaling. When a single processor is transformed into multicore processor, it faces a lot of hazards to confine the circuits into single chip. To emphasize the importance of multicore architecture, this paper provides a comprehensive survey on multicore architectures designs, constraints, and practical issues.

Keywords System-on-chip · Multicore · Cache coherence · Memory system · Parallel and distributed architecture · Tile-based architecture

Abbreviations

ALU	Arithmetic logic unit
CMOS	Complementary metal oxide semiconductor
CPU	Central processing unit
DSP	Digital signal processing
FPOA	Field-programmable object array
FPU	Floating point unit
I/O	Input/output

R. Radhika (✉)
St. Peter's College of Engineering and Technology, Chennai, Tamil Nadu, India

N. Anusha
S.A. Engineering College, Chennai, Tamil Nadu, India

R. Manimegalai
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

IC	Integrated circuit
IMAP	Integrated memory array processor
KB	Kilobyte
MAC	Multiply-accumulate
NoC	Network-on-chip
PE	Processing element
RAM	Random access memory
RISC	Reduced instruction set computer
ROM	Read-only memory
SoC	System-on-chip
VLIW	Very long instruction word

76.1 Introduction

A multicore processor is an integrated circuit (IC) in which two or more processors are clued together for better performance. Multicore architecture has been envisioned for enhanced performance, reduced execution time, and simultaneous processing of multiple different tasks. In general, an individual separate core consists of a central processing unit (CPU), a register set, cache, bus, memory, and I/O ports. In single-core architecture, the abovementioned units are integrated onto a single integrated circuit die. Typical single-core architecture is shown in Fig. 76.1a. Multiple single cores are basically combination of single-core systems with minimum interaction between them, but they can act independently. Though there are tools available for supporting multiple cores, load balancing is one of the major drawbacks in multiple cores. Necessary hardware should be designed to support load balancing by providing I/O queues in network interfaces. Multiple single cores are illustrated in Fig. 76.1b. In a multicore processor, individual cores cooperate with each other in order to gain improved performance. Each core consists of CPU, registers, and cache memory of its own. Multicore architecture reduced overhead, improves cache coherency, and has better possibilities for load balancing. The cons in multicore processors include the following: (i) porting from single core is complicated and (ii) cache coherency issues exist between the cores leading to dependencies among various tasks. A typical multicore architecture on a single integrated circuit (IC) die is shown in Fig. 76.1c.

76.2 Challenges and Issues in Multicore Architecture

Certain tribulations and argument move ahead when the multiple cores are implemented and set upon a single chip. Some of the issues and challenges in multicore architectures are briefly described in the subsequent subsections.

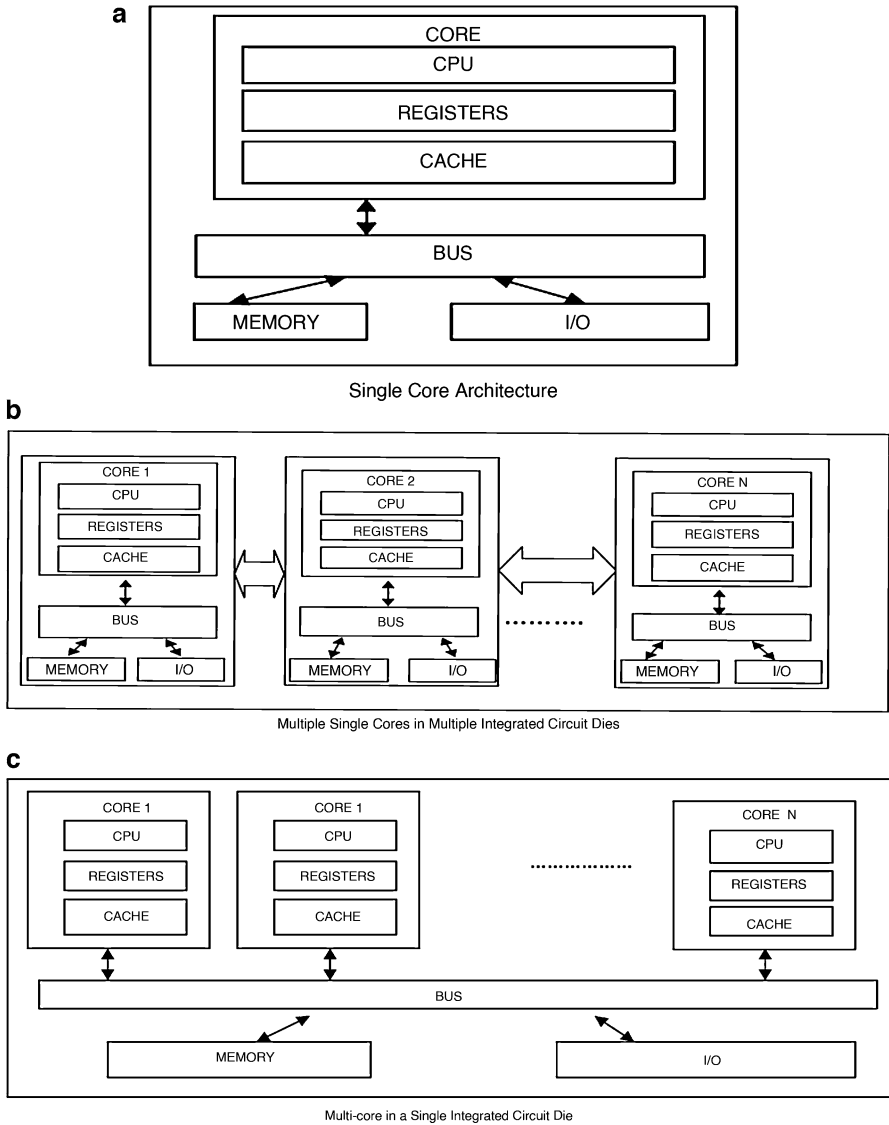


Fig. 76.1 Single core, multiple single core, and multicore architectures. (a) Single-core architecture. (b) Multiple single cores in multiple integrated circuit dies. (c) Multicore in a single integrated circuit die

Cache Coherence Important issue in a multicore environment is cache coherence. Different levels of caches such as L1 and L2 across the chip. Snoopy protocol and directory-based protocol are used to handle the cache coherence issues. Snoopy protocol takes the support of a number of states and these states can determine the

type of values in the cache to be updated. The other protocol is directory based which is highly scalable. This protocol can easily be used on a random network and thus it can be easily scaled to multiple processors or cores [1].

Power and Temperature Implementation of two or more cores on a single chip without any kind of design optimization makes the chip to consume double the power. It also generates larger amount of heat in the chip. In the current trend, every design can be integrated with a power control unit which can shut down the cores that remain unutilized, thus, restricting the power consumption. By restricting the quantity of hot spots over the chip, the generation of large amount of heat can be controlled [1].

Memory System When more numbers of cores are placed on a single chip, more instructions are executed within the chip, and therefore, more memory requests are issued to handle large amount of data. It is necessary to develop an enhanced memory system [1]. In most computers, there must be a delay between memory controllers signal to the memory module and data access from RAM [2]. In order to minimize memory access time, the access should be stimulated as soon as the response of the memory reference becomes available.

Delay in Getting the Data Multicore system is used to run the single-threaded applications and programs. Thus, a proper replacement policy should be used to remove all the cache entries that have been processed by other cores [1]. With the addition of more and more cores in the processor, this problem gets more deep and difficult.

76.3 Design Models of Multicore Processors

The stored program concept was introduced in the late 1940s by John von Neumann.

The design models of multicore processors belong to one of the following architectures such as parallel architecture, distributed architecture, tile-based architecture, programmable device-based architecture, and network-on-chip architecture.

A. Parallel Architecture-Based Design Model

In the middle of the 1990s, UC Berkeley developed a processor named PADDI-2 for digital signal processing (DSP) applications [3]. Four PEs are combined to form a cluster. PADDI-2 has 48 PEs in 12 clusters. Imagine Processor was released by Stanford during late 1990s for stream style applications in media processing [4]. Khailany et al. (2002) have designed and fabricated a prototype for Imagine Processor which has an array of 48 floating point arithmetic units organized as 8 clusters [5]. Each cluster contains six processors. The processors belong to wide Very Long Instruction Word (VLIW) processor with 9.7 KB of local register file and 128 KB of global register file. These register files and memory play important role in Imagine Processor. Most of the computations can be finished in the registers, thus

improving enhancement power as well as energy efficiency. Each lane runs at 800 MHz at 1.0 V and occupies 3.2mm^2 and provides the peak performance of 128GMACs with power consumption of about 10 W. Stanford had proposed Hydra in the late 1990s [6]. Global bus is used to connect the chip with the couple of Reduced Instruction Set Computer (RISC) processors and L2 cache. UT Austin developed TRIPS in the early 2000s. TRIPS is used to provide configurability for parallelism. A 16-bit wide processor is available in each core of TRIPS. In 2001, integrated memory array processor (IMAP) is the fourth-generation linear processor arrays [7]. It integrates 128 VLIW processing elements (PEs) and a RISC control processor which provides single instruction stream for the processor array. It operates at 100 MHz. Picochip has developed Picoarray for multicore DSP in 2002 [8]. This integrates 250–300 individual DSP cores onto a single die. Each of these cores is a single-bit processor with Harvard architecture. The programming model allows each processor to be coded independently and then to communicate with each other. Sixteen clusters are available in Wavescalar architecture in which four clusters contain a traditional L1 data cache [9]. Shared bus is used for intra cluster communication. Ambric's programming model assumes that a modern chip can integrate with multiple cores [10]. Simple software object can run exclusively on its own processor. More complex objects may run on multiple processor cores. Locally, each small cluster of processors in Ambric's programming model runs on its own clock speed and is dictated by its software workload. Globally, however, the processors do not match in time to a single clock signal. IBM cell processor [11] contains 32-KB L1 instruction and data cache and 512-KB L2 unified cache. High-performance asynchronous data transfer, along with parallel data processing on the PPE and SPEs, eliminates the need to explicitly interleave data processing and transfer at program level.

B. Distributed Architecture-Based Design Model

Intelasys has unveiled its SEAFORTH40C18 in 2008 [12]. It is a 40-core processor designed for embedded wireless, portable, and distributed data processing applications. It has an array of 40 fully functional CPUs operating asynchronously. Each of the cores acts as complete computer with its own ROM, RAM, and interprocessor communication. The core size design is 0.13mm^2 and consumes power of 7 mW.

C. Tile-Based Architecture Design Model

The first tile-based multicore processor called RAW was designed at MIT during the late 1990s [13]. Each core contains 32-KB instruction memory, 32-KB data cache, and 64-KB static memory. RAW is composed of a set of interconnected tiles, each tile comprising instruction, switch instruction, data memory, an arithmetic logic unit (ALU), floating point unit (FPU), registers, a dynamic router, and a programmable switch. It verifies high economic constraints and controls interprocessor communication by suitable software. One more tile-based architecture named Smart Memories was proposed by Stanford in the early 2000s [14]. The clock and voltage of each core in the synchrosclar architecture can be scaled down in order to

reduce power consumption [15]. Synchrosalar tiles are based on Blackfin two-way VLIW digital signal processor. These tiles are arranged in columns.

D. Programmable Device-Based Design Model

A reconfigurable digital signal processor called Pleiades Maia for wireless baseband digital signal processing was developed by UC Berkeley in the late 1990s [16]. This architecture consists of an array of computational units of different modules with an embedded microprocessor. Due to this nature, Pleiades Maia architecture achieves high-energy efficiency and also provides flexibility while it limits its application domain. In the early 2000s, CMU manufacturers developed the architecture called Pipe Rench [17]. Pipe Rench is organized by 16 stripes and each stripe consists of 16 processing elements (PEs). Dynamically reconfigurable data path is one of the key features available in Pipe Rench. This feature is used to execute numerically intensive applications efficiently. Xelerated's design relies on patented dataflow architecture in 2000 [18]. Like an ASIC, it is optimized in relation to the flow of packets instead of flow of instructions through a RISC core. As a result, processing speed is predictable rather than variable. Because Xelerated's Ethernet switches are programmable, carrier could reprogram themselves to adapt to new standard as they develop. Domain-specific reconfigurable architecture RaPiD is proposed by the University of Washington in 2001 [19]. It has an array of computational units that vary in complexity. The size and complexity of the computational units are termed as granularity. The computational unit in this architecture can be configured to execute a variety of DSP applications. In 2016, Kilo Core (KC)-256 offers massively parallel processing and great flexibility via dynamic reconfiguration of processing functions, while providing unparalleled computing energy. Ultra-high performance and very low-power consumption are needed for a wide range of compute-intensive applications (Rapport KC256 2016) and [20]. This chip contains 256 independent processing elements (PEs) that can be interconnected arbitrarily and clocked up to 100 MHz to perform 25 billion operations per second effectively while consuming less than 500 mW.

E. Network-on-Chip Architecture-Based Design Model.

Intel 80 core is a network-on-chip (NoC) architecture containing 80 tiles arranged as a 10×8 2D mesh network [21]. Field-programmable object array (FPOA) in 2007 resides in core and periphery area. Core area contains three types of core objects, namely, arithmetic logic unit (ALU), multiply-accumulate (MAC), and register files. Periphery area includes internal and external RAM, general-purpose I/O, transmit interface, and receive interface. FAUST chip architecture in 2015 integrates 20 asynchronous NoC nodes, 23 synchronous units including an ARM946 core, embedded memories, various IP blocks, reconfigurable data path engines, and 1 clock management unit to generate the 23 distinct unit clocks [22]. It has 4–16 cores that operate at 2.2 GHz and a peak performance of 40 Gbps of application processing with extremely low-power consumption of 8 W. Its size is about 37.5 mm.

76.4 Conclusion

Today's multicore processor designs usually come with quad-core and 8-core processors. This lets us to know that the multicore architecture is promising in the future. Intel has duplicated a chip with 80 cores on board already, just to see if it was possible. There is also the introduction of the 22-nm microprocessor using high-K metal gates where K is a dielectric constant that is used in characterizing capacitors. These high-K metal gates have been stated to reduce leakage power in the processor. The leakage power has just recently become a big factor due to the small transistor scaling. Software, unfortunately, has not been making such fast advances. Researchers have generated parallel code through the compiler by employing software pipelining, but the hunt for creating parallel algorithms and programming remains the same. Even for multicores today, programming multiple cores is a difficult task. Once this breakthrough occurs, it will be a big benefit for both software and hardware. Software engineers will have an easier job programming on multicore processors, and hardware designers will have better utilization of their designs.

References

1. Mohanty RP, Turuk AK, Sahoo B (2012) Analysing the performance of multi-core architecture. 1st International conference on Computing, Communication and Sensor Networks-CCSN, pp 1–6
2. Das B, Mahato AK, Khan AK (2013) Architecture and implementation issues of multicore processors and caching—a survey. *Int J Res Eng Technol* 2(2):78–82
3. Keung A, Rabaey JM (1995) A 2.4 GOPS data-driven reconfigurable multiprocessor IC for DSP. *IEEE International conference on solid-state circuits*, pp 108–110
4. Rixner S, Dally WJ, Kapasi UJ, Khailany B, Lopez Laguns A, Mattson P, Owens JD (1998) A bandwidth-efficient architecture for media processing. *IEEE international symposium on microarchitecture (MICRO)*, pp 3–13
5. Khailany B, Dally WJ, Chang A, Kapasi UJ, Namkoong J, Towles B (2002) VLSI design and verification of the imagine processor. *IEEE International conference on computer design*, pp 289–294
6. Hammond L, Hubbert B, Siu M, Prabhu M, Chen M, Olukotun K (2000) The stanford hydra CMP. *IEEE J Micro Archit* 20(2):71–84
7. Kyo S, Koga T, Okazaki S, Uchida R, Yoshimoto S, Kuroda I (2003) A 51.2 GOPS scalable video recognition processor for intelligent cruise control based on a linear array of 128 4-way VLIW processing elements. *IEEE International conference on solid-state circuits*, pp 48–49
8. Carlstrom J, Nordmark G, Roos J, Boden T, Svensson L, Westlund P (2004) A 40Gb/s network processor with RISC dataflow architecture. *IEEE International conference on Solid-state circuits*, pp 60–61
9. Swanson S, Michelson K, Schwerin A, Oskin M (2003) Wavescalar. *IEEE international symposium On microarchitecture (MICRO)*, pp 291–302
10. Jones AM, Butts M (2006) TeraOPS hardware: a new massively-parallel MIMD computing fabric integrated circuit. In: *Proceedings of IEEE hotchips symposium*, pp 59–66

11. Pham D, Asano S, Bolliger M, Day MN, Hofstee HP, Johns C, Kahle J, Kameyama A, Keaty J, Masubuchi Y, Riley M, Shippy D, Stasiak D, Suzuki M, Wang M, Warnock J, Weitzel S, Wendel D, Yamazaki T, Yazawa K (2005) The design and implementation of a first generation cell processor. *IEEE international conference on solid-state circuits*, pp 184–185
12. Intellasis Seaforth-24B Embedded array processor, technical report. <http://www.intellasis.net/>
13. Waingold E, Taylor M, Srikrishna D, Sarkar V, Lee W, Lee V, Kim J, Frank M, Finch P, Barua R, Babb J, Amarasinghe S, Agarwal A (1997) Baring it all to software: raw machines. *IEEE J Comp Soc* 30(9):86–93
14. Mai K, Paaske T, Jayasena N, Ho R, Dally WJ, Horowitz M (2000) Smart memories: a modular reconfigurable architecture. *International symposium on computer architecture*, pp 161–171
15. Oliver J, Rao R, Franklin D, Chong FT, Akella V (2005) Synchrosalar: evaluation of an embedded, multi-core architecture for media applications. *J Embed Syst Spec Issue Multi-Core Archit*:1–16
16. Zhang H, Prabhu V, George V, Wan M, Benes M, Abnous A, Rabaey JM (2000) A1-V heterogeneous reconfigurable DSP IC for wireless baseband digital signal processing. *IEEE J Solid State Circuits (JSSC)* 35(11):1697–1704
17. Schmit H, Whelihan D, Moe M, Levine B, Taylor RR (2002) PipeRench: a virtualized programmable datapath in 0.18 micron technology. *IEEE Custom Integrated Circuits Conference (CICC)*, pp 63–66
18. Baines R, Pulley D (2003) A Total cost approach to evaluating different reconfigurable architectures for baseband processing in wireless receivers. *IEEE Commun Mag* 41(1):105–113
19. Cronquist DC, Franklin P, Fisher C, Figuerar M, Ebeling C (1999) Architecture design of reconfigurable pipelined datapaths, *Conference on Advanced Research in VLSI*, pp 23–40
20. Xanthopoulos T, Chandrakasan AP (2000) A low-power DCT core using adaptive bit width and arithmetic activity exploiting signal correlations and Quantization. *IEEE J Solid State Circuit* 35(5):740–750
21. Vangal S, Howard J, Ruhl G, Dighe S, Wilson H, Tschanz J, Finan D, Lyer P, Singh A, Jacob T, Jain S, Venkataraman S, Hoskote Y, Borkar N (2007) An 80-tile 1.28 TFLOPS network on-chip in 65nm CMOS. *IEEE International conference on solid-state circuits*, pp 98–99
22. Kasapaki E, Schoeberl M, Sørensen RB, Müller C, Goossens K, Sparso J (2015) Argo: a real-time network-on-chip architecture with an efficient GALs implementation. *IEEE transactions on very large scale integration systems*, pp 1–14

Chapter 77

Blockchain-based e-Voting as a Service



R. S. Shyam Prakash and G. R. Karpagam

Abstract Blockchain is one of the most innovative things in recent times in computing. It is a verifiable and immutable ledger of records. The rise of cryptocurrencies has brought them more focus. It is being made clear that the applications of blockchain are not only restricted to cryptocurrencies but also can extend up to the way we use our smartphones, the structure of security systems, banking systems, health monitoring systems, even e-governance, etc. In this paper, an attempt is made to show the importance of blockchain in e-governance, through e-voting. The impact of blockchain along with its synergy with e-voting has been discussed in this paper, which can be used for further investigations.

Keywords Blockchain · e-governance · e-voting · Distributed ledger

Abbreviations

DDoS Distributed denial of service
MD5 Message Digest 5
SHA Secure Hash Algorithm

77.1 Introduction

In the recent years, much attention in computing technologies has been shifted towards blockchain – the distributed, secure ledger technology. This paper attempts to explain the synergy between blockchain technology and e-voting. Blockchain is a peer-to-peer system which is made up of several individual nodes in a network.

R. S. Shyam Prakash (✉) · G. R. Karpagam
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

The blockchain database is a simple file that just stores transaction data, which is of minimal size and is encrypted. Grouping of blocks is done based on timestamp. The transactions are irreversible and all the nodes can have access to the information, but cannot delete or change a transaction. Blockchain found its most important application in bitcoin which has experienced a surge of value, over 700 percentage, on the basis of year over year. In this paper, we try to explore the use of blockchain in e-governance, by discussing about blockchain-based e-voting system.

77.2 Blockchain Fundamentals

Blockchain eliminates the requirement of a trusted third party for validation of transactions. Rather, a trusted environment is created which comprises multiple peer nodes, which are hashed together using cryptographic algorithms such as SHA, MD5, etc. The reason for which blockchain is more powerful than traditional databases is that it is decentralized. Data is submitted on approval of all the other peers and can be viewed by them later as an option of inspection or review.

77.2.1 Blockchain Applications

(a) Finance

The use of blockchain has been extensively applied for financial transaction, leading to the invention of cryptocurrency. Creation of each block requires the values of previous block, which is supplied by the blockchain miner. The process of addition of new block inside the chain is called mining. The legitimacy of each block along with its validity is checked from the blockchain, which on failure removes the block from the transaction pool. Blocks are added only when a greater number of peers accept the inclusion. As the process of mining has been very resource intensive, it has made the attacker very difficult to validate an invalid transaction.

(b) Healthcare

The need for managing healthcare-related data is increased day by day with the increase in patient numbers. The process of storing and processing information becomes harder for hospitals and clinics with the increase of data volume over the years. In drug dealer's application, for example, the immutability and time stamp nature of blockchains enable the feature of fraudulent detection in between drug

dealers. Also, when a drug is produced and is moved from the manufacturer to retailer, the recording of operational data done on a blockchain makes it easy to verify the path of the drug and determine all its associated links at any time.

(c) e-Governance

e-Governance is the use of information technologies for improvement of efficiency, effectiveness, transparency, and accountability of governance in any country. It is moving the services offered to citizen online in a view of cost reduction, which promotes economic development, increases transparency, and also improves delivery of services and administration. The objectives of e-governance can be listed as:

77.2.1.1 Cost Reduction

The processing costs are reduced when services are deployed online, when compared with the manual handling of governmental operations. Attainment of efficiency is also made possible by well running of processes which enables faster decision-making.

77.2.1.2 Enhancement of Accountability and Transparency

As stated in [2], e-governance helps in the improvement of the transparency of decision-making processes by making information widely accessible such as expression of budgets and expenditure statements and key decisions and allowing the online tracking of applications over the Web by the public.

77.2.1.3 Service Delivery Improvement and Facilitation of e-Society

The traditional government service delivery process lacks transparency and is time consuming, which ultimately leads to dissatisfaction of citizens. When the government services are made online, reduction of bureaucracy and improvement of quality of services happen automatically with the benefit of using information technology in several sectors. The technological skills required for e-governance administration support the development of new training courses and modules in educational institutions willing to supply the required skills and capabilities to the society. Also, on implementation, blockchain has the potential to extend the capability of many applications in number of domains as stated in [4].

77.3 Characteristics of Voting

77.3.1 Accessibility

Convenience in access and equality should be ensured in an electronic voting system. Accessibility refers to both physical access with understanding in use of the technologies and working of an e-voting system. The casting of vote should be free from any special restrictions or skills and familiarity in the voting machine.

77.3.2 Accuracy

An electronic voting system should ensure that all valid voters vote only once and the valid votes casted are counted accurately. The traditional voting systems have shown chances of compromise in the Ballot, where casted votes are stored. Thus a well-designed e-voting system will be perceived as more accurate than traditional voting systems.

77.3.3 Privacy

An e-voting system should ensure that the link between voter and ballot can never be made. Political activities also need to ensure the protection of privacy. The threats to privacy of voters can cause negative response in voters.

77.3.4 Mobility

Since the early 2000s, the participation of voters has increased with the flexibility and mobility of e-voting. It also encourages voters to vote from their own location, which eliminates the restrictions of availability.

77.4 Synergy Between Blockchain and e-Voting

The use of blockchain in voting for elections can impose the above said properties of the e-voting. Blockchain does not have any proprietary elements in it, and it allows any citizen or agency to audit the functionality of the application. Many elections

still are conducted using centralized Web-based server or even in paper-based system, which leaves huge holes for security breaches. Transparency, being one of its traditional assets, allows the votes to be followed up, counted, and correlated by various sources while still keeping the privacy of the voters intact.

77.4.1 Confidentiality and Integrity

The blockchain-based e-voting scheme is secure cryptographically by exhibiting the property of blocks being tamper resistant, which ultimately satisfies the properties such as confidentiality and integrity of voting process. In addition to it, nonrepudiation is also taken care. This ensures any voter or candidate that the votes casted by them are authentic, and contents are maintained with integrity. Thus it is highly impossible for anyone to change or duplicate other person's vote. Another way of ensuring the integrity of electronic voting machines is independent verification of software and certification.

77.4.2 Auditability and Accuracy

Various auditing mechanisms are available to detect any possible compromise of security in the construction of the vote ballots and in the tellers operation during the tabulation phase. The ballots have unique hashing values and IDs allowing auditing of ballots before the election, but they require random testing of ballots before the election can be done to ensure their correct construction. The ballot audit would make it possible to confirm that the ballot is encrypted block with candidate identifiers. Verifiable Tallies need to be generated after election. In addition, it has to be made sure that the all the verification and processes are entirely independent.

77.4.3 Privacy

The voter's privacy property is satisfied through the use of implied hashing techniques used to build the decentralized application. The enforced privacy of a blockchain-based e-voting system ensures that a voter marks his ballot privately. The random candidate ordering in the ballot provides secrecy of vote and adds up to the voter's privacy. Long-term privacy is satisfied by use of hashing process of the user account details.

77.4.4 Authentication

The blockchain-based e-voting system satisfies authentication property by only allowing voters satisfying the voter's requirements to vote, and only voters listed in the registration phases are allowed to vote. The voters are authenticated using their ID and private/public keys. The system also adheres to the policy that no voter should be able to vote more than once; therefore, a voter gets access to only one ballot, at most once.

77.4.5 Mobility of Ballots and Voters

The blockchain-based e-voting scheme has got the property to satisfy voter mobility through making it available online. A number of ballots could be generated for each district or zone as per the requirement and are stored on a secure server. These blockchain-based ballots can be used on demand, on site, and in any station for polling.

77.5 Agent-Based Model and Its Need

An agent-based model (ABM) is a computational model in which the actions and interactions are done by autonomous systems. The authoritative decision of deciding whether to perform an action or not is taken based on the information obtained. The main difference between a traditional software program and an agent is that the latter are more autonomous, are proactive, and are highly scalable. The need for accuracy and independent behavior of the proposed system arises the necessity for incorporation of agent-based model. The various agents present along with their roles are depicted in Fig. 77.1 and are described as follows.

77.5.1 Registration Agent

Registration process is the initiator of the voting system, as it starts with feeding of information about the candidates contesting in the election and the voters willing to vote. It is carried out by the registration agent in which the user is allowed to enter the credentials, along with supporting proofs verifying his citizenship and authenticity and making him eligible to participate in the process of election, if they are valid.

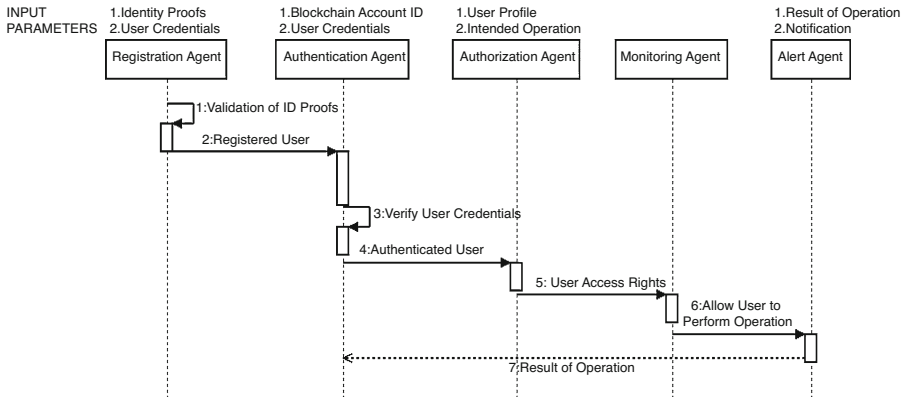


Fig. 77.1 Agents and their roles

77.5.2 Authentication Agent

The authentication agent is secure enough to verify the identity of anyone who connects to the voting system. When an ID is added to the blockchain, the identification issuing module binds a public key and then transfers the private key to the user. This allows only the user to sign a signature which is verified against the public key stored in the blockchain. This identification of user serves as decentralized source of authentication.

77.5.3 Authorization Agent

Authorization is the process of determining whether a user has authority to access the requested service, i.e., to vote. It is closely coupled with authentication agent as the user must be authenticated in order to get authorized. Authorization is very much required because in the e-voting system, the election commission authorities, the voters, and the candidates contesting would have different access rights based on their role. Thus, it is essential to have a policy to allow access to only resources which they are abide to.

77.5.4 Monitoring Agent

Monitoring agent observes and reports on the overall functionality of the system. They ensure that the access rights are not violated, which is a main objective of the system, as confidentiality is an essential quality to be maintained. Any abnormalities in the system are reported to the reporting agent, to which control is transferred.

77.5.5 Alert Agent

The declaration of events or announcement of results is done through the alert agent. It receives input from various other agents present in the system. The alert is transferred to the concerned individual or group of people through the desired media. In fact, alert agent is one of the most important agents in the system. The interaction between agents has been depicted through sequence diagram in Fig. 77.1.

77.6 The Need for Blockchain-Based e-Voting and Its Advantages

Conduction of fair election is essential in democracy, as the trust of people toward it is declining. The interest of people towards functioning of electoral systems are increasing as stated in [1]. With the current voting systems being weak in security, they can be manipulated [3]. The cost of conduction of such traditional elections is also high, and also blockchain-based system has several substantial benefits. The resource utilization results from simulation of election environment have been summarized using table and graphs in Figs. 77.2a and 77.2b, with reference to [5].

The overall working of the system with the messages between agents has been depicted in Fig. 77.3. The system also primarily eliminates the need of user visiting the voting booth, by making it accessible by smartphones or computers with Internet access. Users who cannot afford to have the access to smartphone or Internet-enabled computers can vote from the temporary booth setup at different locations. There are a number of network attacks possible with voting systems, affecting operations, as listed in Table 77.1. The incorporation of blockchain makes the

Fig. 77.2a CPU utilization %

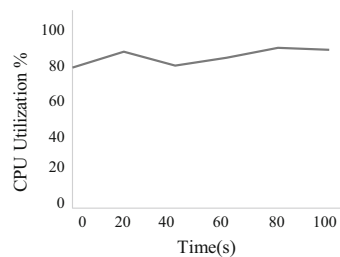
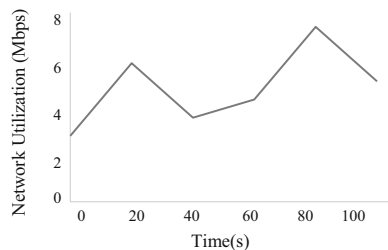


Fig. 77.2b Network utilization (in Mbps)



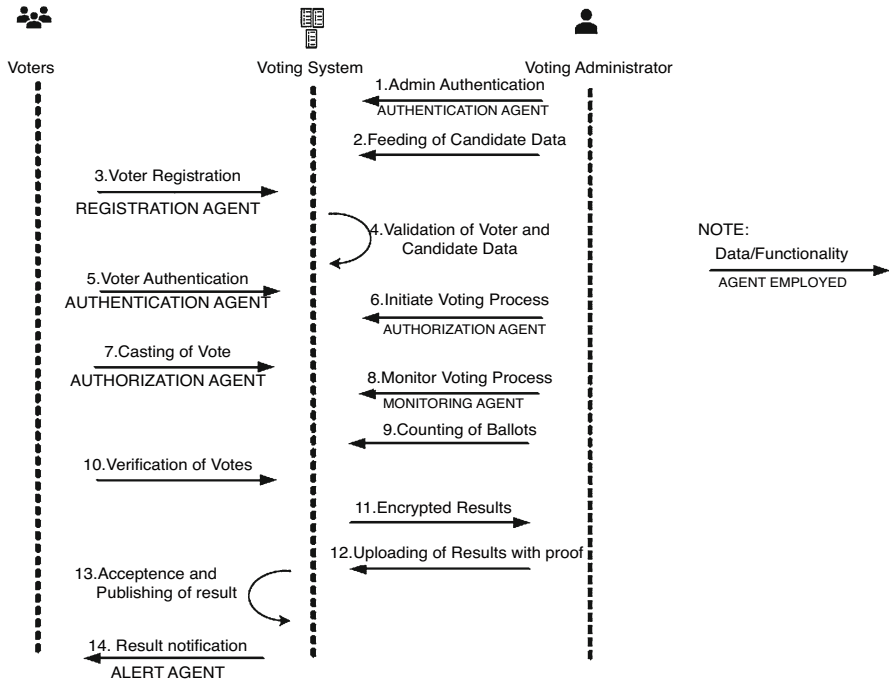


Fig. 77.3 Overall working of the system

Table 77.1 Security threat and processes affected

Security threat	Method	Processes under hazard in e-voting
Network threat	Distributed denial-of-service (DDoS) attack	Access to e-voting system, performing intended operations
	Man-in-the-middle attack	Accurate casting of vote, retrieval of candidates, and user profile

Table 77.2 Comparison of execution of system on Intel i3 and i5 processor

Number of candidates	Execution time	
	i3 CPU	i5 CPU
1	1.52 s	1 s
5	3.49 s	2.46 s
15	6.12 s	4.97 s
25	13.45 s	9.47 s

proposed system resistant from two of the most common attacks listed below, which similar systems experience. Performance results shown in Table 77.2 make it evident that higher throughput can be achieved by increasing the number of processor cores, on which the application is deployed.

77.6.1 Distributed Denial-of-Service (DDoS) Attack

DDoS attack is feasible to happen in traditional systems, as the control and data are maintained in a centralized fashion. A DDoS attack on Ethereum blockchain occurs when a person attempts to exploit the network's resources in a way that the other users are unable to do any other operations.

Protective Feature

As Ethereum transactions require ether for every transaction to be done, it is nearly impossible to have an account wallet which is full of ethers forever. All the gas in any delegated account (wallet) gets empty at some point of time. Whenever the account wallet is empty, the operation loop is terminated, and thus, the network is protected from DDoS. Also it is highly impossible for the attacker to take control of numerous systems at once. This ensures and protects the Ethereum blockchain against DDoS attacks.

77.6.2 Man-in-the-Middle Attack

Man-in-the-middle attack is a technique in which an attacker probably modifies the transmitted data between two communicating parties. In the traditional methods, the attackers have a possibility to intrude between systems and can modify the voting information, which ultimately leads to compromise of security.

Protective Feature

The blockchain-based system has strong resistance to this man in the middle as the voting data and user credentials are strongly encrypted with public and private keys. Hashing algorithms are used to make the system more free from forged addition of customized attacker data. The ballot to which the votes casted are recorded is also encrypted strongly, owing to the design and structure of blockchain.

77.7 Conclusion

With the growth of blockchain-based applications and experiments, faith on the long existence of blockchain technology is increasing. It is obvious that the participation of voters in democracy increases with increase in the trust of process, which can be attained by transparency and ease of access, an essential property of the proposed system. Maintenance of privacy of voters has also been taken care in the system which further builds confidence, enabling users to grow their intention to participate in the duty of democracy.

References

1. Wang B, Sun J, He Y, Pang D, Lu N (2018) Large-scale election based on Blockchain. *Procedia Comp Sci* 129:234–237
2. Øines S, Ubacht J, Janssen M (2017) Blockchain in government: benefits and implications of distributed ledger technology for information sharing. *Gov Inf Q* 34(3):355–364
3. Ayed AB (2017) A conceptual secure blockchain-based electronic voting system. *Int J Netw Secur Its Appl (IJNSA)* 9(3):01–09
4. Aras ST, Kulkarni V (2017) Blockchain and its applications – a detailed survey. *Int J Comp Appl* 180(3):0975–8887
5. Tuan T, Dinhz A, Wangz J, Chenx G, Liuz R et al BLOCKBENCH – A framework for analyzing private blockchains

Chapter 78

Optimum Resource Allocation Techniques for Enhancing Quality of Service Parameters in Cloud Environment



M. Kandan and R. Manimegalai

Abstract Cloud computing offers multitenancy with countless services and follows pay-and-use strategy. In this paper Quality of Service (QoS) parameters such as energy consumption and response time are considered for resource allocation. Customer satisfaction can be fulfilled by improving the QoS. Multi-Agent-based Dynamic Resource Allocation (MADRA) strategy, a multistage framework using QoS-based Resource Allocation (QRA) algorithm, and Artificial Immune System-Directed Acyclic Graph (AIS-DAG) model are proposed for optimum resource allocation. The performance of proposed approaches using the CloudSim toolkit is analyzed and compared. The experimental results show that proposed approach has high potential for the improvement in QoS and scalability by concentrating on resource and request validation.

Keywords Cloud computing · Request validation · Resource allocation

Abbreviations

AI	Artificial Intelligence
AIS-DAG	Artificial Immune System-Directed Acyclic Graph
CCR	Communication Computational Ratio
DC	Data center
DEA	Differential Evolution Algorithm
DVFS	Dynamic voltage and Frequency Scaling

M. Kandan (✉)

Sasi Institute of Technology and Engineering, Tadepalligudem, West Godavari District, Andhra Pradesh, India

R. Manimegalai

Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_78

DVS	Dynamic voltage scaling
ERA	Efficient resource allocation
GA	Genetic Algorithm
GC	Green Cloud
ICSA	Improved Clonal Selection Algorithms
IDEA	Improved Differential Evolution Algorithm
IoT	Internet of Things
LA	Local Agents
MACS	Multiagent Cloud System
MADRA	Multi-Agent-based Dynamic Resource Allocation
QoS	Quality of Service
QRA	QoS-based Resource Allocation
QS	Quality score
SLA	Service Level Agreement
VM	Virtual machine

78.1 Introduction

Cloud is one of the fastest-growing computing environments to improve and enhance communication in the world of Internet of Things (IoT). Cloud is a set of hardware, software, network, storage, and interface that empower the delivery of computing as a service. Depending on the users' demand, Cloud services include delivery of software, infrastructure, and storage over the Internet [1]. Similarly, various types of resources are available in the Cloud that can be accessed only by authorized users. The two different ways of accessing Cloud resources are free and Pay-and-Use mode.

Resource allocation is the technology that utilizes the shared pool of computing resources in the network to facilitate the execution of complicated tasks in Cloud [2]. Resource allocation needs to consider many factors such as load balancing, completion time, response time, and energy consumption. In Cloud, resources need to be allocated not only to satisfy Quality of Service (QoS) but also to reduce response time and energy consumption [3]. Cloud computing meets a lot of issues and problems continuously while managing the requests and resources. Due to the increasing number of user requests and resources deployed in the Cloud, it is difficult to manage the requests because it takes more amount of time and energy to provide the appropriate resources effectively. And also scalability issues occur due to the fluctuation in the user requirements for resources in the Cloud.

78.2 Related Works

This section gives a brief review about the various existing resource allocation techniques which mainly consider the response time and energy efficiency of resources in Cloud. The comprehensive survey of resource allocation strategies discusses about various optimal allocations techniques used for resource allocation [4]. Improved Differential Evolution Algorithm (IDEA) combines the Taguchi method and a Differential Evolution Algorithm (DEA). The DEA has a global exploration capability on macro-space. The Taguchi method is used to exploit the better individuals on micro-space to be potential offspring. The result improves completion time, response time, and cost [5]. In Multiagent Cloud System (MACS), Global Agent (GA) dispatches user requests to multiple Local Agents (Las), while the LAs make optimal resource management and Genetic Algorithm (GA) for searching the solution of global request. The proposed MACS approach reduces energy consumption and improves scalability [6]. A framework is used to improve the scalability and resource allocation by validating user request and resources in the server. The user request and resource are validated by local manager and global manager [7].

One of the major concerns in Cloud environment is power consumption. In order to reduce power in Cloud, resource scheduling algorithms are categorized into three main categories based on the relevant works in literature [8] which are Dynamic Voltage Scaling (DVS) oriented, Artificial Intelligence (AI) oriented, and Hybrid. In modern portable devices, DVS is a novel technology which is used to optimize the supply voltage and frequency of the processor(s) to minimize energy utilization. To apply DVS in the scheduling procedure, the tasks can be defined as both in a directed acyclic graph [9] and in a non-graphical form. In [9] task graph is defined with nodes representing tasks and their execution time and the extremity representing the communication cost between tasks. Efficient Resource Allocation (ERA) approaches to reduce power consumption by considering the Data Center (DC) and the load. It optimizes the load and schedule and estimates the power consumption using a linear and quadratic technique [10]. FUGE combines fuzzy logic theory and GAs while considering execution time and cost. They modify the standard GA to improve its performance by devising a fuzzy-based steady-state GA. Their suggested architecture includes four layers, namely, fabric layer, unified resource layer, a platform layer, and application layer [11]. Improved Clonal Selection Algorithms (ICSA) select the best resource and allocate resource to user requests with minimal time by cloning and mutation. The ICSA approach reduces makespan, response time, energy consumption, and cost. In this approach compilation time and response time can be further reduced by improving the communication between the nodes [12]. An elastic power-aware resource provisioning approach

(ePower) is applied for heterogeneous workloads in self-sustainable data centers that completely rely on renewable energy. The goal of this work is to control the system power consumption with respect to green power supply [13].

78.3 Proposed Works

This section discusses about various proposed resource allocation techniques such as a Multi-Agent-based Dynamic Resource Allocation, a multistage framework using QoS-based Resource Allocation, and Artificial Immune System-Directed Acyclic Graph to improve QoS parameters such as response time and energy consumption. Finally, result of each technique has been analyzed and compared.

78.3.1 Multi-Agent-Based Dynamic Resource Allocation

This work presents a Multi-Agent-based Dynamic Resource Allocation (MADRA) strategy for Distributed Cloud. The agents are used for improving the QoS with reference to resource allocation. In this proposed work, the entire processes are decomposed into several sub-processes and are assigned to agents in the Cloud server. Green navigator agent is used to navigate the request with appropriate resources. The service analyzer agent interprets and analyzes the incoming service requirement of users. It decides whether to accept or to reject the request based on the status of load and energy in the virtual machine (VM) manager and energy monitor. User profile agent collects the details of activities and characteristics of Cloud user. An essential Cloud requests are granted special privileges and prioritized over other users. The QoS parameter monitoring agent analyzes the user requests and matches with the resources. If there is a mismatch, then those resources are omitted. It calculates quality score (QS) for each matched resource in terms of time, energy, and availability and is represented in Eq. (78.1). Then, the resources are ranked based on quality score. Service scheduling agent selects the highest quality score of the resource and is allocated for the appropriate incoming request which is represented in Eq. (78.2). Pricing agent decides about the cost for services based on the time, demand, and supply required for the resources [14].

$$QS(i) = QS_i(\text{Time}) + QS_i(\text{Energy}) + QS_i(\text{Availability}) \quad i = 1, 2, \dots, n \quad (78.1)$$

$$\text{Maximum_QS} = \max (QS(i)) \quad i = 1, 2, \dots, n \quad (78.2)$$

78.3.2 *Multistage Framework for Improving QoS in Resource Allocation*

This work discusses about an efficient resource allocation technique with improved QoS to get economic benefits. The proposed framework has three different layers such as server layer, communication layer, and resource layer integrated with three different agents such as server agent, communication agent, and data center agent. There are three components used in the entire Cloud processes, namely, (i) Cloud server, (ii) server manager, and (iii) virtual machines. Cloud server comprises server manager and virtual machines to handle the requests by utilizing the server virtualization method. The server manager is used to update and show the present processors information to the incoming clients, whereas virtual machines are allocated to handle the incoming requests and data server processed response is forwarded to the server manager. All the requests in the request queue are investigated by agents to find out if it is a valid request or not. If the requests are not valid, then they are eliminated; else the request is moved to the server layer by agents [15].

Procedure QRA ()

```
{
1. If all the requested processors are present in the nearest server manager, then the
server manager allocates the processors to client request.
2. If only a few processors are present in the server manager, then the request is
divided into the number of sub-tasks and available processors are allotted.
3. If the server manager is already allocated and it is in the end stage of the process,
then the client is assigned to wait state (for a minimum time constraint) and then
the requests are loaded into the data server.
4. If the client did not receive any response according to their request within the
stipulated time interval, then search the next nearest server manager by agents.
}
```

All the clients submit their request to the server manager and then the data server loads the requests to its server manager. Server manager processes the requests according to the following QRA procedure. The calculated QoS parameters such as bandwidth, memory storage, and CPU are compared with the threshold values of the resources using Eq. (78.3). If all the computed QoS parameters of the resources meet the following constraints, then the resource is allocated to a client request. QRA procedure notices the power consumption and the resource scheduling to reduce energy and response time.

$$\text{If } (QoS(R_i) \geq Th_i, 1 \text{ else } 0 \quad (78.3)$$

78.3.3 Artificial Immune System-Directed Acyclic Graph Model-Based Resource Allocation

In this work, an artificial immune system algorithm is used to allocate an accurate resource to the user request. The communication model directed acyclic graph is utilized to improve the communication and make the Cloud server to take the accurate decision for resource allocation [16]. The proposed Artificial Immune System-Directed Acyclic Graph Model analyzes the user requests and assigns the accurate resource. It also improves the QoS parameters such as time and energy [17]. In AIS-DAG, multiobjective optimization such as energy and time is optimized together to allocate best resource. They are represented using Eqs. (78.4) and (78.5). The energy is consumed for completing the job J_j on r_i at k Dynamic Voltage and Frequency Scaling (DVFS) level.

$$E_i = \gamma \times f \times \left\{ \sum_{j \in J(i), k \in L(j)} \left([v_k(i)]_j^2 \times CT(i, j) \right) + v_{\min}(i) \times f_{\min}(i) \times Idle_i + \lambda \right\} \quad (78.4)$$

$$M_s = \min \left\{ CT_{ij} \mid J_j \in J, j = 1, 2, \dots, n \text{ and } R_i \in R, i = 1, 2, \dots, m \right\} \quad (78.5)$$

AIS-DAG-Based Resource Allocation The AIS-DAG model is applied for resource allocation to contract with the optimization problem where the affinity function (fitness) is contracted with Energy (E_i) efficiency, Makespan (M_s). The affinity function in AIS algorithm is defined in Eq. (78.6). The main objective function of AIS-DAG model is to allocate the best resource according to optimum values such as less energy, less time, available resource, and taking less cost for processing the job in the resource.

$$\text{aff}(x) = e^{\min E_i + \min M_s} \quad (78.6)$$

Algorithm: Artificial_Immune_System

- {
1. Generate the random population antibodies $P(0)$; the total size of P is S .
2. Calculate the affinity of antibody $P(k)$.
3. Select the $R\%$ of higher-affinity antibodies as the population $P_l(k)$.
4. Clone the individuals $P_l(k)$ and get $P_c(k)$.
5. Apply mutation on the population $P_l(k)$ to form the new mutated population $P'(k)$.

6. Compare the affinity of $P'(k)$ with $P_l(k)$; replace the lesser affinity with the higher affinity.
 7. Select the population $P'(k)$, and obtain the new population generation as $P(k + 1) = P_C(k) \cup P'(k)$, $k = k + 1$.
 8. Repeat the above steps 2–4 until $K > X_G$ reached [stopping criteria] or the reached P_m .
- } Where S is the population size, P populations, XG maximum generation, Pm mutation probability, K iteration operation, and R half of the antibodies.

The main objective function of AIS-DAG model is to allocate the best resource according to optimum values such as less energy, less time, available resource, and taking less cost for processing the job in the resource.

78.4 Experimental Results and Discussion

In this section, the efficiency of the proposed approaches is investigated by simulating its procedure in Green Cloud (GC) Simulator software tool. GC is based on NS-2 simulation platform. It is able to make Cloud elements such as data centers, switches, servers, and network links. GC supports three-tier architecture-based application simulations. The simulation parameters are network size, number of nodes, communication frequency, and Communication Computational Ratio (CCR). Proposed work proves that AIS-DAG is the best in terms of response time as shown in Fig. 78.1. In terms of relevancy, while comparing requests incoming vs. response time, the AIS-DAG model improves the response time. While comparing with MADRA and QRA approaches [14, 15], the AIS-DAG model [17] has shown 20% and 10.56% reduction in response time.

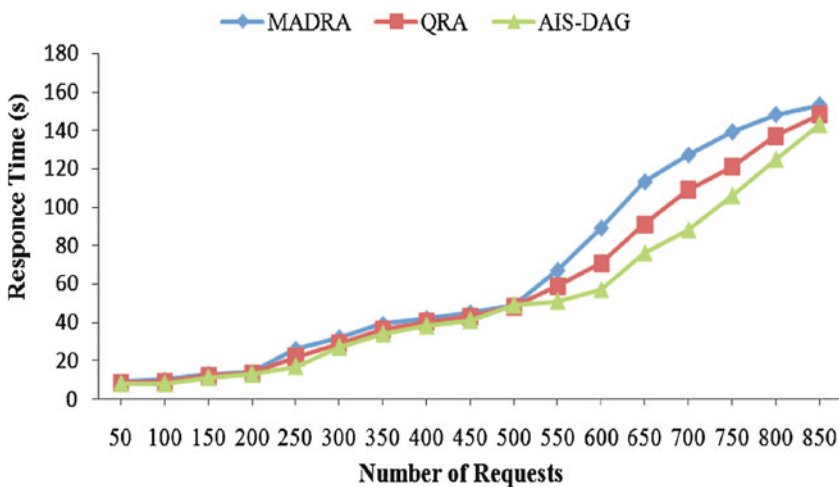


Fig. 78.1 Comparison based on response time

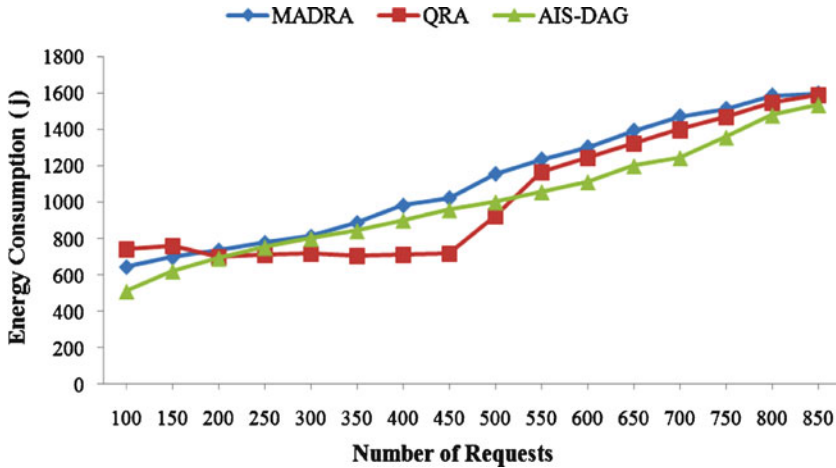


Fig. 78.2 Comparison based on energy consumption

In this section, the energy consumption is calculated. The amount of energy required depends on the number of processes, size, and type of the resource. The AIS-DAG model reduces energy consumption by 10.02% and 2.33% when compared with MADRA and QRA techniques, respectively. From the simulation, it is clear that the AIS-DAG model takes less amount of energy than other proposed techniques. The energy consumed by all the proposed techniques is shown in Fig. 78.2.

78.5 Conclusion

In this paper, an efficient resource allocation technique has been proposed for Cloud resource allocation with low cost, faster response time, and optimum energy consumption. All the proposed strategies were analyzed and compared. The experimental results show that AIS-DAG model improves the response time which demonstrates high potential in improving the energy efficiency of the data centers in the Cloud. Also, it can effectively fulfill the Service Level Agreement (SLA) requested by the users. The energy efficient resource selection and allocation technique AIS-DAG proposed in this paper can be further extended to mobile Cloud which is a trending technology for energy saving and increased battery life.

References

1. Armbrust M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A, Stoica I, Zaharia M (2010) A view of cloud computing. *Commun ACM* 53(4):50–58
2. Huang CJ, Guan CT, Chen HM, Wang YW, Chang SC, Li CY, Weng CH (2013) An adaptive resource management scheme in cloud computing. *Eng Appl Artif Intell* 26(1):382–389
3. Rodero-Merino L, Vaquero L, Gil V, Galan F, Fontan J, Montero R, Llorente I (2010) From infrastructure delivery to service management in clouds. *Futur Gener Comput Syst* 26(8):1226–1240
4. Kandan M, Manimegalai R (2015) Strategies for resource allocation in cloud computing: a review. *Int J Appl Eng Res* 10(76), 1–76, 10
5. Tsai JT, Fang JC, Chou JH (2013) Optimized task scheduling and resource allocation on cloud computing environment using improved differential evolution algorithm. *Comput Oper Res* 40(12):3045–3055
6. Sun P, Dai Y, Qiu X (2017) Optimal scheduling and management on correlating reliability, performance, and energy consumption for multiagent cloud systems. *IEEE Trans Reliab* 66(2):547–558
7. Kandan M, Manimegalai R (2015) A framework for effective resource allocation in a distributed cloud environment. *Int J Appl Eng Res* 10(87):493–498
8. Abolfazli S, Sanaei Z, Gani A, Xia F, Yang LT (2014) Rich mobile applications: genesis, taxonomy, and open issues. *J Netw Comput Appl* 40:345–362
9. Meikang Q, Yang LT, Shao Z, Sha EHM (2010) Dynamic and leakage energy minimization with soft real-time loop scheduling and voltage assignment. *IEEE Trans Very Large Scale Integr Syst* 18(3):501–504
10. Vakiliinia S, Heidarpour B, Cheriet M (2016) Energy efficient resource allocation in cloud computing environments. *IEEE Access* 4:8544–8557
11. Mohammad S, Saeed J, Saeid A, Nicola C (2015) FUGE: a joint meta-heuristic approach to cloud job scheduling algorithm using fuzzy theory and a genetic method. *Cluster Comput Springer* 18(2):829–844
12. Shu W, Wang W, Wang Y (2014) A novel energy-efficient resource allocation algorithm based on immune clonal optimization for green cloud computing. *EURASIP J Wirel Commun Netw* 2014:2–9
13. Cheng D, Rao J, Jiang C, Zhou X (2016) Elastic power-aware resource provisioning of heterogeneous workloads in self-sustainable datacenters. *IEEE Trans Comput* 65(2):508–521
14. Kandan M, Manimegalai R (2015) Multi agent based dynamic resource allocation in cloud environment for improving quality of service. *Aust J Basic Appl Sci* 9(27):340–347
15. Kandan M, Manimegalai R (2017) QRA: a multi-stage framework for improving QoS in resource allocation. *J Adv Res Dyn Contr Syst* 9(5):131–141
16. Kliazovich D, Johnatan EP, Andrei T, Pascal B, Samee UK, Zomaya AY (2015) CA-DAG: modeling communication-aware applications for scheduling in cloud computing. *J Grid Comput* 14(1):23–39
17. Kandan M, Manimegalai R (2016) AIS-DAG: artificial immune system for directed acyclic graphs model based fair resource allocation for heterogeneous cloud computing. *Asian J Inf Technol* 15(19):3673–3686

Chapter 79

MIPGIOT: Monitoring and Improving the Productivity in Garment Unit Using IOT



V. G. Prabhu and R. Manimegalai

Abstract Enhanced communication and increase in radio frequency improve the growth of emerging Internet of Things (IoT) applications. Textile manufacturers are always trying to improve the production and quality of the garments to sustain in the enormous competitive market. They perform enormous number of operations at different spots by the operators to determine the sustainability and profitability. All these activities need to be performed in synchronized and timely manner to achieve desired productivity. Smart/interactive textiles are one of the methods of deploying smart materials in textile sectors. Smart materials appear to think and memory to revert back to their original state and also communicate with the master system. Machines are deployed with various smart devices and they are made to interact with the main master system. These data are collected in a storage device, and they are transferred to master system for monitoring the user efficiently, by determining the machine's ideal state. The master system deploys various data analytic algorithm for measuring the performance of the system when it receives the status of the machine. Depending upon the prior values, a model can be developed so that it can be used to reveal estimate time to complete the given task.

Keywords Internet of Things · Interactive · Master system · Ideal state · Garment manufacturing unit

Abbreviations

DMAIC	Define, Measure, Analyze, Improve, Control
IoT	Internet of Things
ISO	Indian Standard Organization

V. G. Prabhu · R. Manimegalai (✉)
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu,
India
e-mail: drmm@psgitech.ac.in

QMS	Quality management system
QSS	Quality Six Sigma
SAM	Standard Allowed Minute

79.1 Introduction

The Internet has emerged as a new era in enhancing the digital and physical devices over the Internet which is known as Internet of Things (IoT). These devices differ from the traditional devices operated in the Internet. Smart devices are objects that have an *identity* in the network, which can be localized, can process and communicate data, and can interact among them to the surrounding environment. Image processing techniques are deployed to determine the texture, and this information processing and communication will support decision-making, improve situational awareness, increase operational efficiency of processes, and pave the way for new revolutionary applications in the widest range of fields. Textile is one of the significant segments in world trade and product manufacturing. The technology development and the growth of competition in the society makes us to make a new solution that helps to improve the applications in the textile field.

79.2 Literature Survey

Jadhav et al. [1] have discussed about a time-based study on determining the productivity of the garment unit by determining the efficiency of the garment. Jadhav et al. propose an idea of achieving the target by performing the required activities in synchronized and timely manner to achieve desired productivity. This work elaborates on the scope for reduction in completing the given task and proposes some considerable ideas in productivity. Jadhav et al. explain about the importance of the skills required for an individual line performing complex task which will increase the efficiency and productivity. This work provides an idea about efficient tool in determining and improving the productivity rate in garment unit. Khantun [2] observes the time required for determining the scope for reduction and productivity improvement in garment industry. The concluded parameters that are to be concentrated for improving productivity are the target time in order sheet and time required for supply of pieces.

Pranjali et al. [3] discuss about textile industries investments amount in maintaining the company's output by deploying several managers in governing the workers and equipping the plant with work system by integrated system of men, machine, & material and also measuring the operation time. Pranjali et al. find some issue in the existing methodology of increasing productivity in garment manufacturing, which is done by unsupervised managers evaluating the performance of the users at different checkpoints that reduces the cost for larger automated machines.

Basak [4] describes about industrial engineering and various other methods that are combined with industrial engineering and other operational management systems that are considered by Pranjali.

Sachidanand et al. [5] propose an idea in inferring about a decentralized network in textile sector for adopting new technologies which will not only manage supplies but also control production and enhance productivity. The proposed work describes about QMS/ISO practices and Six Sigma with their own benefits so that when they are implemented, they are complement with each other and the overall productivity is improved. Bewoor [6] considered a quality management system named DMAIC, and the proposed technique is compared to existing method DMAIC as a part of QMS. Sachindanand et al. [5] address a problem as there is no prior research or study made in integrating Six Sigma and QMS. This work is assumed to be a successful progress and coined as QSS which would be an effective approach in increasing the productivity in textile sector. Sachindanand et al. describe about the need to adopting a new improvement technology that not only manages supplies but also controls production and enhances productivity.

Yuchen et al. [7] have done a survey on Internet of Things that are everywhere in our daily life and the benefits that can come of huge risks of privacy loss and security issues. Yuchen et al. differentiate IoT and the traditional devices with Internet as the absence of human role and describe IoT devices which can create information about individual's behaviors, take action, and analyze them. Granjal [8] et al. prove the importance of IoT applications which are offering great benefit in human's life and specify the importance in security protection and person's privacy. Aumasson [9] et al. survey presented the importance in security and privacy issues in IoT applications and systems. The authors also discuss the limitations in IoT devices in battery and computing resources. Yuchen et al. advise the IoT manufacturing industry to work closely with security and demand the standardization organizations to tackle newly emerged threats as well as to develop robust and strong security standards for IoT devices and systems.

79.3 MIPGIOT: The Proposed Methodology for Measuring and Improving the Productivity in Garment Unit

79.3.1 MIPGIOT: Proposed Idea

MIPGIOT is the proposed method for Monitoring and Improving the Productivity in Garment Unit using IOT. Every garment manufacturing plant has manual intervention for measuring the productivity metrics. This project aims to automate the monitoring process using IoT-based solution. The images of all types of garments manufactured are stored in a database. When the garments are getting stitched, the captured images are matched against the stored images to identify the expected number of pieces to be manufactured within a time period. The data captured through sensors are analyzed to determine the number of piece manufactured per

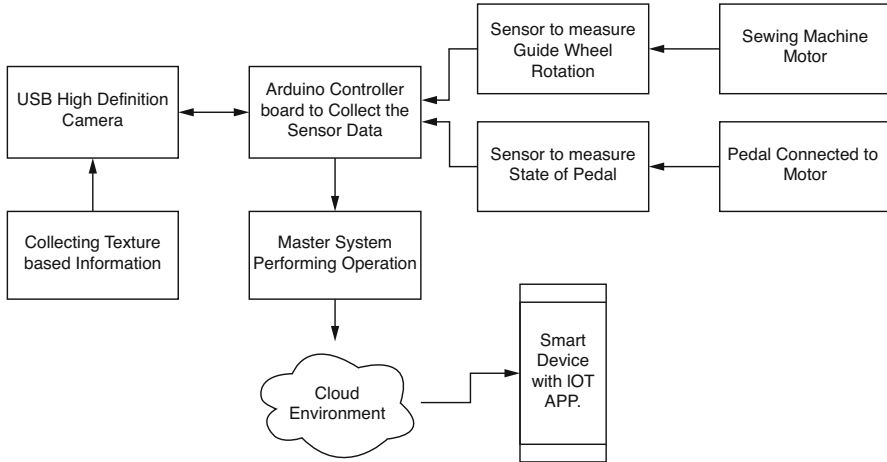


Fig. 79.1 Block diagram of the proposed MIPGIOT

day and the idle time of the sewing machine. These productivity metrics are communicated to all stakeholders (CEO, managers, and supervisors) of the garment manufacturing unit.

Every sewing machine is deployed with smart device to measure time required to complete the given task as shown in Fig. 79.1. Cameras are used to determine the texture and size of the material to evaluate the required time for completing the task. These values are secured through an encryption algorithm and the cipher data are transmitted to a master system. Machine collects the data for performance calculation by some data analytic algorithm, and the performance result is uploaded to private cloud which can be accessed only by authorized persons.

79.4 Implementation of the Proposed Method

79.4.1 MIPGIOT: Implementation Process

The proposed method works on the method of monitoring the line efficiency through integration of various sensors, actuators, and cameras to acquire the data from the sewing machine and determine the efficiency by evaluating the data obtained from the Internet of Things. These determined values are uploaded to cloud via Internet so that stakeholders can view and instruct the supervisor of corresponding department to take necessary action over the users to maintain the productivity of the garment unit. The method proposed acts as a live monitoring technology in maintaining the productivity and user performance of the garment unit. Master machine compares the obtained data from the sewing machine and with the captured image. Simple image processing technique is applied to equate the image with the existing database. Some predetermined values are measured and uploaded to a database for

comparison. The obtained results are compared with the existing database and observed that faster and better results are obtained. Data from sensors are encrypted by using an encryption algorithm by a microcontroller. Images are directly transferred to the master machine for processing. The state of the pedal is monitored to ensure no false values are measured from RPM sensor. Microcontrollers trigger the camera and transmit the data. It provides an updated outcome of the overall performance of the company based on individual user performance without much of human involvement.

79.4.2 MIPGIOT: Results Achieved

Data are encrypted by the controller in the sewing machine and these data are transferred to the master machine. The data are classified based on the MAC ID of the controller encrypted along with the data which is used to differentiate the data from various other systems. The images captured are transferred with an individual ID given for each camera to differentiate the image from each system shown in Fig. 79.2. The time required for this process takes around 3 s, and the data are continuously obtained and processed by the master machine. The complete process

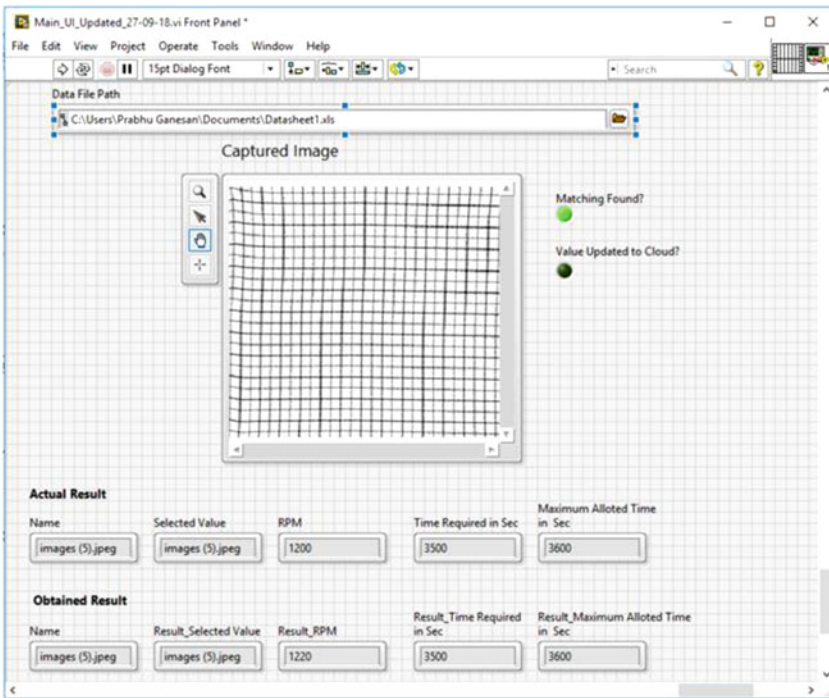


Fig. 79.2 Output obtained in MIPGIOT

is automated so that no manual interventions are required so that the overall errors are minimized and logs of the data are stored for future analysis.

79.5 Conclusion

The proposed method is used to determine the performance of garment unit without any human intervention. The microcontrollers were programmed using Arduino IDE with a RSA encryption algorithm [10], and the image processing is developed in LabVIEW. The performance of the RSA algorithm comparatively improves the key generation speed and highly secured the protection of data. In this method information processing and communication will support decision-making, improve situational awareness, increase operational efficiency of processes, and pave the way for new revolutionary applications in the widest range of fields. As RSA algorithm is used for encryption, the encrypted data cannot be found easily. Images are to be compared with existing database for comparison. It can be extended to weaving and embroidery machines efficiency and exiting loading of products. It can also be developed to automate monitoring of factory operations and equipment maintenance.

References

1. Jadhav SS, Sharma GS, Daberao AM, Gulhane SS (2017) Improving productivity of garment industry with time study. *Int J Textile Eng Process* 3(3):1–6. ISSN 2395-3578
2. Khatun M (2013) Application of industrial engineering technique for better productivity in garments production. *Int J Sci, Environ Technol* 2(6):1361–1369
3. Chandurkar P, Kakde M, Bhadane A (2015) Improve the productivity with help of industrial engineering techniques. *Int J Textile Eng Process*, ISSN: 2395–3578, 1(4):35–41
4. Avizit B (2015) Supply chain management in garments industry. *Global J Manage Bus Res* 14 (11):23–28
5. More SS, Pawar DMS (2013) Performance improvement of textile sector by implementing quality six sigma. *Int J Appl Innov Eng Manage*:4847, ISSN 2319 –
6. Bewoor AK, Pawar MS Developing integrated model of six-sigma methodology and quality management system for improving quality productivity and competitiveness, 12th annual conference of society of operations management, I.I.T.-Kanpur, Kanpur, India
7. Yang Y, Wu L, Yin G, Li L, Zhao H (2017) IEEE a survey on security and privacy issues in internet-of-things. *IEEE Internet Things J* 4(5):1272–1283
8. Granjal J, Monteiro E, Silva JS A secure interconnection model for IPv6 enabled wireless sensor networks. In: *Proceedings of IFIP wireless days, Venice, Italy, Oct 2010*, pp 1–6
9. Aumasson J-P (2017) The impact of quantum computing on cryptography. *Comp Fraud Secur* 2017(6):8–11
10. Stallings W (2006) *Cryptography and network security: principles and practices*, 4th edn. Pearson, Upper Saddle River

Chapter 80

Text and Audio Transfer Using LI-FI Technology



S. Sabareeswaran, G. Madumitha, and M. Shruthi

Abstract Wired or wireless network has become the essential way to use the Internet and complete people's task these days. As the number of people accessing the wireless network increases, it gets complicated to rely on an incorrupt signal due to the jammed airwaves. Though Wi-Fi delivers a speed up to 1300 Megabits per second, it is not adequate to hold all the users. To redress this drawback of Wi-Fi, a new concept is being introduced called as Li-Fi, which provides a speed up to 224 Gigabits per second. This is very analogous to Wi-Fi. The major disparity is that visible light is used by Li-Fi and radio frequency is used by Wi-Fi to transmit data. In this project an Android application is created to transfer binary data from one device to another device through Li-Fi, whereas for audio transfer, a hardware setup is implemented using LED, solar cell, and speaker.

Keywords Li-Fi · Electromagnetic spectrum · Wi-Fi · Visible light communication

Abbreviations

Li-Fi	Light Fidelity
Wi-Fi	Wireless Fidelity
OOK	On-Off Keying
VLC	Visible Light Communication
RF	Radio Frequency
LED	Light Emitting Diode
FOOS	Fress and Open Source Software

S. Sabareeswaran · G. Madumitha · M. Shruthi (✉)
Department of Information Technology, P.S.G. College of Technology, Coimbatore, Tamil Nadu, India

80.1 Introduction

This chapter gives a brief description about Li-Fi and the problems faced by the developers and the users in adapting to this technology. It also compares the electromagnetic spectrum of both radio waves and visible light.

80.1.1 Light Fidelity

People rely more on the Internet for their day-to-day activities. It is absurd these days to even think of not being associated with the “Internet.” Especially when one wants to transfer data instantly and accurately, low Internet speed could be vexatious. In 2011, Professor Harald Haas, during his TED talk, proposed the concept of “data through illumination.” Haas used fiber optics to transmit data through LED light bulbs. Using Li-Fi, one can access the Internet easily within the range of an LED beam.

80.1.2 Problem Definition

Light modulation is an existing concept, but adapting to the technology will take some more time because replacing Wi-Fi with Li-Fi is not an easy task since people have been adapted to Wi-Fi for years. People have become more comfortable with Wi-Fi technology so evolving to a world replacing Wi-Fi for people takes some

- To Developers – It is quite challenging for developers to make an existing device to adapt to a new technology with some improved software.
- To Users – Switching over to a new technology from an existing technology.

80.1.3 Objective

The sole purpose of this project is to provide an efficient, comparatively cheap, riskless, controlled, and swift data transfer technique which can be an alternative for the conventional technique, i.e., Wi-Fi. Simultaneously, Li-Fi also lets us to adapt to more efficient light source, i.e., LED. The prime objective is to design an application that transmits data be it text, audio, or video using Li-Fi technology for coping with the limited bandwidth problem we face in RF (radio frequency). One of the benefits of practicing Li-Fi over Wi-Fi is that it avoids radiation produced by Wi-Fi.

80.2 Literature Review

Harald Haas showed a demo of streaming a HD video through Li-Fi at a rate of 10Mbps which is faster than average broadband connections.

Most people adopt practicing wireless fidelity devices because it will be useful for considerable amount of radio frequency (for a say up to 5G hertz) to provide Wi-Fi Internet connectivity encompassing the area of workplace, school, home, and public places [1]. Wi-Fi's bandwidth is extended to a larger extent it may not be the capable way to add advanced facilities like global position system application and gesture detection [2].

There is a section of electromagnetic spectrum that does not enormously use the visible light [3] now called as Li-Fi. Contrarily, they bid an exclusively new prototype in wireless technologies regarding the speed of communication, usability, flexibility, and reliability. Visible light is the potential key to the overall wireless spectrum scarcity. VLC uses visible light between 400 THz and 375 THz as a data communication medium as similar to optical carrier for illumination and transmission of data. By varying the flickering rate of light (i.e., modulating the LD light source), the data is encoded to generate a new data.

Light is considered to be the chief part of our life and yet has no side effects. Furthermore, it is abundant in nature. Transmitting data starts from an LED bulb, by constantly changing the current at excessively large speed which cannot be detected by human eye [4]. The data is accepted by a silicon photodiode. The capacity of switching velocity is acquired by LED which helps them to inflect depending on the flow of bits. This takes place in a parallel flow so that data transmission is large, concurrently [5]. One of the greatest advantages is that Wi-Fi does not work under the water's surface, but Li-Fi does, as light can be used with much ease [6]. Li-Fi will be smaller than any of the systems we have up to date. It will be low-priced with the rate of titanic data transmission [7].

The whole new spectrum of possibilities [8] is almost 10,000 times more in size to the radio waves spectrum. Visible light which is a mandatory part of an infrastructure apart from being innocuous to vision is also densely available and easily attainable.

80.2.1 Drawbacks of Wi-Fi

- Wireless networking which is similar to any other radio frequency is subject to a wide heterogeneity of interferences as well as to complex/compound propagation effects that are beyond the command of the network administrator.
- Private information could be misused or stolen using the access points when transmitted from Wi-Fi users.
- Wireless networks (1–54 Mbps) are slower than that of wired networks (100 Mbps to Gbps).

- Illegal 2.4GHz spectrum is used by Wi-Fi. This spectrum is cramped with other devices like microwave ovens, video sender devices, cordless phones, Bluetooth, etc. This causes deterioration in the performance.
- Wi-Fi networks have only a finite spectrum. A home router has an acceptable spectrum of 45 m inside and 90 m outside.

The project focuses on transferring the binary data using Li-Fi technology through Android application and audio transfer using LED and solar cell. The application gets the binary value from the user and sends the values through the LED available in the smartphones, and it is received by another smartphone through the light sensor present in it. In audio transfer audio file is given as input and output is obtained through speaker.

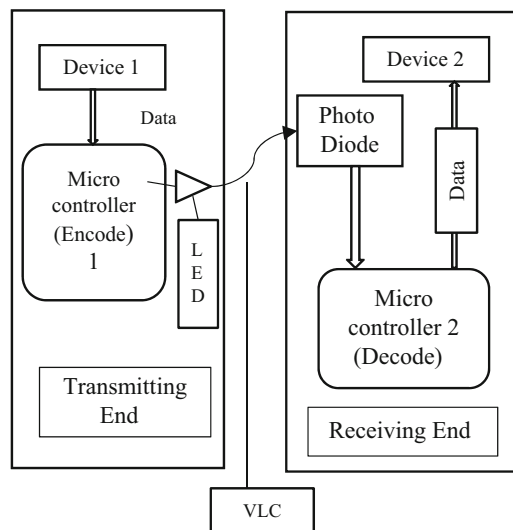
80.3 System Design

The block diagram in Fig. 80.1 shows how the data is passed from one device to another through Li-Fi. It shows the overall working of the application.

80.3.1 Working of Data Transfer

In the transmitter end, a smartphone having Android 6.0 and above is used. In the receiver end, another smartphone having Android 6.0 and above is used for data transfer. This Android application uses inbuilt flashlight (LED) as transmitter and ambient light sensor as a receiver to transmit a data between two Android

Fig. 80.1 Block diagram for data transfer



smartphones. Here photo diode is also called ambient light sensor. Microcontroller in transmitter and receiver is used for convert the signal into data. The communication takes place through VLC between two Android smartphones.

80.3.2 Working of Transmitter

In the transmitter part, the binary data is obtained from the user first and then the value is stored. And depending on the values, the flashlight transmits the data. The working of the transmitter is shown in Fig. 80.2.

80.3.3 Working of Receiver

In the receiver part, the light from the device which is sending the data is captured using the light sensor present in it, and depending on the intensity of the light, the values are obtained and stored. The working of the receiver is shown in Fig. 80.3.

Fig. 80.2 Flowchart of transmitter

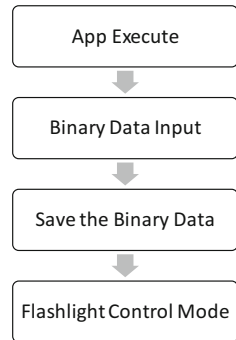
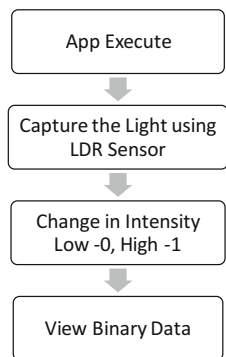


Fig. 80.3 Flowchart of receiver



80.4 Software Implementation

Transfer of binary data through Android application using Li-Fi technology.

80.4.1 Modules Description

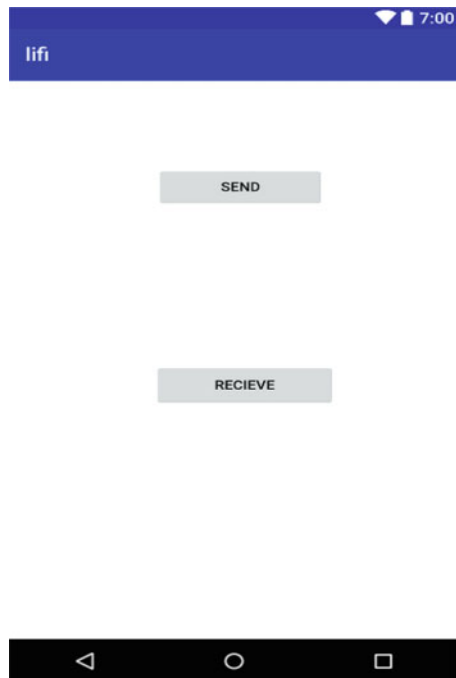
For every product that is being developed, it is very important to divide the modules before developing, because it paves the way for a parallel development, then finally combining it by integration. Likewise, the modules for this product include:

- Home page.
- Data transmission.
- Data reception.

80.4.2 Home Page

This application allows the user to do both send and receive the data in the same application itself. It consists of two buttons one for sending the data and other for receiving the data. The index page of the application is shown in Fig. 80.4.

Fig. 80.4 Index page of the application



80.4.3 Transmitting of Binary Data

The application receives the binary data entered in Edit Text from the user and the user recognizes the length and saves binary data for the length in array. With SetFlashMode method, it runs “torch,” and if the value is 1, the mode is “off” and for the value 0, turning on/off flashlight by OOK (on-off keying) of level detection method. The data transmission happens in binary digits of 1 and 0 as shown in Fig. 80.5.

80.4.4 Receiving the Data

This application uses inbuilt Light Sensor in the receiving Android smartphones to receive the data sent by the sender and displays the data in binary digits as shown in Fig. 80.6.

Fig. 80.5 Data transmission page

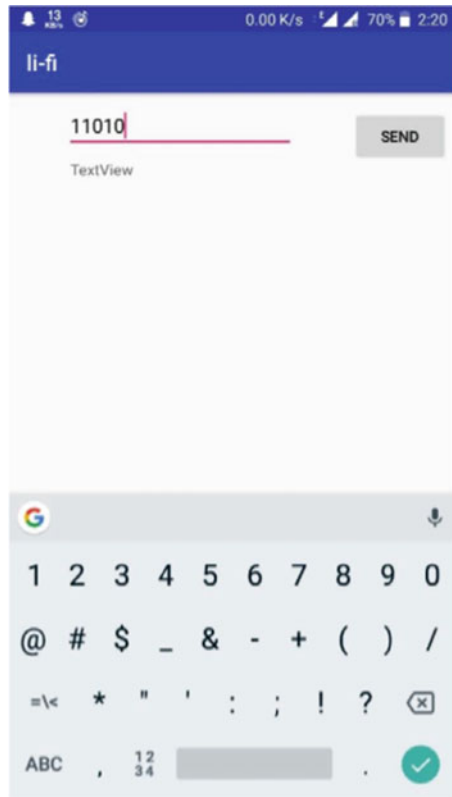
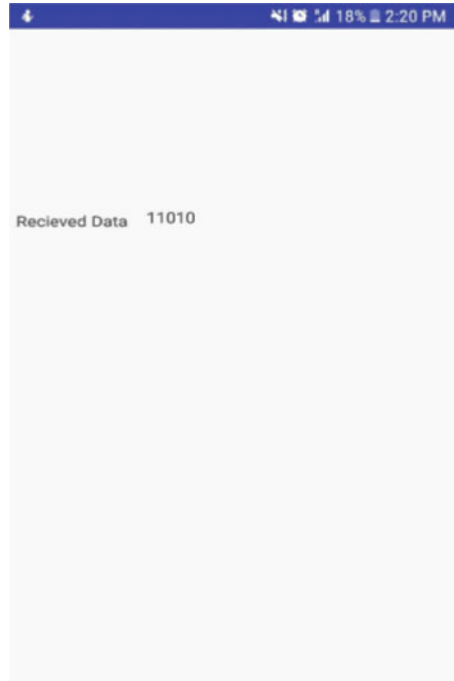


Fig. 80.6 Data reception page



80.5 Audio Streaming Through Li-Fi

80.5.1 Working of Audio Transfer

- (a) *Input*: Input consists of analog signal, which is obtained from the audio output of the mobile phone, laptop, or any other musical instruments. The signal gained is of a low-voltage level which is insufficient to drive an LED, so in order to drive the LEDs, we have to amplify the signal using amplifiers.
- (b) *Output*: The demodulated audible signal is transmitted from speaker to its final destination. Hence, the audience can listen to the message that has been transmitted from the source. The working of audio transfer is shown in Fig. 80.7.

For audio streaming, the components needed are 9-volt battery, smartphone, or any other audio playing device, AUX cable, solar cell, speaker, LED, and connecting wires.

The connections are shown in Fig. 80.8.

First the Aux cable from the smartphone is connected to the LED. Aux cable contains three wires, positive terminal, negative terminal, and neutral. Leave the neutral wire and connect positive of Aux to positive of LED and do the same for negative terminal. The music played from the mobile phone is sound energy which is converted into electrical signal by the inbuilt microcontroller of the smartphone

Fig. 80.7 Block diagram of audio transfer

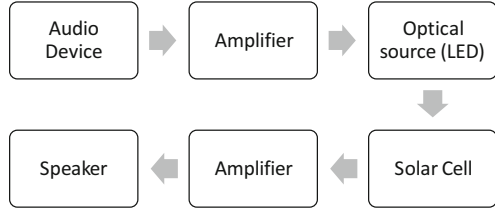
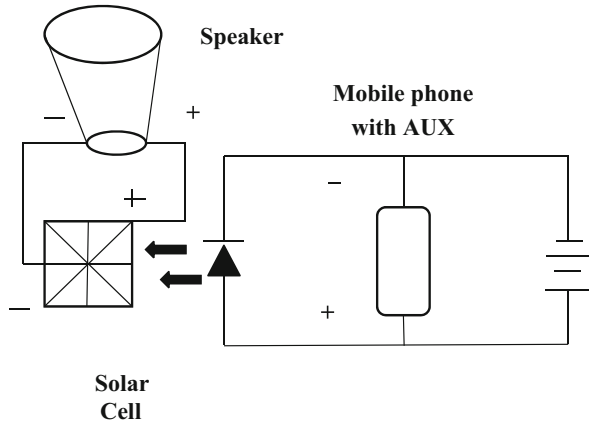


Fig. 80.8 Circuit diagram for audio streaming hardware setup



through Aux cable. Now connect the batteries’ respective terminals to the already connected setup. Now the LED will glow when music is played in the smartphone (electrical energy is converted into light energy). Now solar cell along with speaker setup is kept behind the LED to receive that light energy. Now solar cell is responsible for converting the light energy received from LED into electrical energy and sent to the speakers. Finally, the speaker’s inbuilt amplifier helps to convert that electrical energy into sound energy.

80.6 Implementation

80.6.1 Binary Data Transfer Through Li-Fi

(a) *Software Requirements.*

- *Android Operating System:* Android is hype of the world which is known as the operating system that was invented for the smartphones and tablets by the most reputed company known as Google. It is a modified part of the Linux kernel which is a FOSS (Fress and Open Source Software). Predominantly it was made for the mobile phones using Android OS, tablets, and laptops.

- *Android Studio SDK*: Android is a FOSS platform for mobile phones, tablets, and computers. Android was put forth by OHA that was initiated by Google and a combination of many other companies. Android apps are mostly designed with JAVA language using the Android SDK. The Android SDK comes with an extensive set of advancement tools.
- *Gradle*: Gradle is a FOSS build a mechanized system that puts forth upon the concepts of Apache (Unix based Web server) that initiates a DSL which is an alternate of the XML form used by Apache for proclaiming the project composition.

(b) *Hardware Requirements*

- Android smartphone with flashlight, light sensor capability, and Android version v6.0 (Marshmallow) or above for better functioning of the application.

80.6.2 Audio Transfer Through Li-Fi

(a) *Hardware Requirements*

- *LED* – A light-emitting diode (LED) is a semiconducting device. It uses the concept of p–n junction diode which gives out light when triggered. When a permissible current is given as an input, the electrons will club together emitting energy like in photons. This emission is known as electroluminescence and this is the process behind an LED.
- *Solar Cell* – A solar cell is otherwise known as a photovoltaic cell. It is a device that transforms the light energy into electrical energy called as the photovoltaic effect. It produces current approximately of 0.5–0.6 volts (Fig. 80.9).

The working of a solar cell requires these terms:

- The light ingesting electron-hole pairs.
- The segregating of current carriers of the positive and negative types.
- The individual abstraction of these transporters to a secondary circuit.

Fig. 80.9 Solar cell

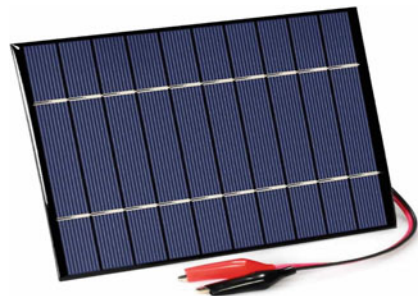


Fig. 80.10 Jbl speaker

- *Speaker:* The demodulated signal is usually of low modulation. So, an amplifier is used to modify it to the arbitrary voltage. The electromagnets will aid to convert the electrical signals into an audible range. In this project JBL speaker is used for implementation as shown in Fig. 80.10.
- Auxiliary Cable
- Connecting Wires
- Battery

80.7 Results and Discussion

Since time is directly proportional to distance, there is a small increase in the data transmission time when the distance is increased between two smartphones.

Parameter measured in air medium between two smartphones through LI-FI technology is listed in Table 80.1 with varying distances. It is clear from the table that when the distance between the transmitter smartphone and receiver smartphone changes, time taken gradually increases for data transmission with light traveling at a stable speed of 300,000,000 m/s. When the distance between two smartphones is varied above 1 m, the LED light gets diverged, and the data cannot be transferred. This can be overcome by introducing powerful LEDs with lesser divergence.

Light can transfer data only in transparent liquid medium (water), vacuum, and air medium. It cannot transfer data in colloidal medium or semi-colloidal medium like oil, kerosene, petrol, curd, etc. This is clearly listed in Table 80.2. Time taken inside the water medium is also the same as air medium since the speed of light is constant. When distance increases, a small time change in data transfer can be encountered.

Table 80.1 Parameter measured in air medium with altered distance

S. No	Parameter measured	
	Distance between two smartphones (in cm)	Time Taken for data transfer (in sec)
1	5	1
2	30	1
3	50	1.0.1
4	100	1.2
5	500	–

Table 80.2 Parameter measured in different mediums

S. No	Parameter measured in different mediums		
	Medium	Data transmission	Time taken
1	Water	Possible	1 s
2	Oil	Not- possible	–
3	Air	Possible	1 s
4	Curd	Not- possible	–
5	Kerosene	Not- possible	–
6	Vacuum	Possible	1 s

80.8 Conclusion and Future Work

This application was run on Android platform mobile devices. The application can turn on/off flashlight by torch “on/off” in set FlashMode method based on OOK (on-off keying) signal of level detection method. Therefore, it is clearly seen that visible light communication among Android platform mobile devices with flashlight and camera can be made by using Android application without having to use any additional devices. The future enhancement for this application would be such that the application will be able to transfer actual data such as documents, images, and videos. There is a better chance for the progression and adoption of Li-Fi technology as light is free to use and omnipresent. A Li-Fi bulb can be used to transfer data if the technology is developed. If this technology is favored, there will be greener, cleaner, secure communication and fulgent environment. The theory of Li-Fi attracts many people as it is license-free and data transmission is faster than that of Wi-Fi. People will use this technology with acceleration, if it expands faster.

References

1. Rani J (2017) Li-Fi (Light Fidelity)-The future technology. *Wireless Commun*
2. Aftab F, Nafees M, Khan U, Ali S (May 2016) Light Fidelity (Li-Fi) based indoor communication system. *Int J Comp Netw Commun* 8:21–31
3. Swami NV (2015) Li-Fi (Light Fidelity) – the changing scenario of wireless communication. *Int J Res Eng Technol* 4:435–438

4. Khaparde SE, Kathane BY (2016) An emerging technology of data transfer through light waves. *Int J Recent Trends Eng Res (IJRTER)* 2:2
5. Shankari RG (2018) Visible light communication using Li-Fi. *Int J Res Appl Sci Eng Technol* 6:3072–3075
6. Vinnarasi A, Aarthy ST (2017) Transmission of data, audio signal and test using Li-Fi. *Int J Pure Appl Math* 117:179–186
7. Chatterjee S, Agarwal S, Nath A (2015) Scope and challenges in light fidelity (Li-Fi) technology in wireless data communication. *Int J Innov Res Adv Eng (IJIRAE)* 2:2349–2163
8. Kolhe AB, Mandavgane RN (2017) Data transmission using Li-Fi system. *Int J Innov Res Sci Eng Technol* 06

Chapter 81

Hyperspectral Image Segmentation Using Evolutionary Multifactorial Spectral Analysis for OMEGA Dataset



Nagarajan Munusamy  and Rashmi P. Karchi 

Abstract Mars region is being imaged with an exceptional combination of spectral and spatial resolution spectrometer using OMEGA instrument. The hyperspectral images of Mars provide spectral range and chemical species with high resolution. This paper presents a novel unsupervised segmentation algorithm named as evolutionary component analysis for remotely sensed hyperspectral image data for material identification in the spatial and spectral information. Sparse multinomial logistic regression (SMLR) algorithm is initially employed to learn the posterior probability distributions from the spatial and spectral information of the images containing class imbalance information to infer the class distribution of the testing hyperspectral data. Evolutionary multifactorial spectral analysis (ESA) helps to characterize noise and extremely mixed pixels with less training set with high training quality and utility with respect to spectral signatures and its spectral changes with less interaction for classification and end-member detection. The proposed segmentation approach based on ESA is investigated and estimated using both real and simulated hyperspectral datasets. ESA is evaluated for the endmember extraction in the mixed pixel revealing up-to-date performance when compared with advanced hyperspectral image classification techniques. The combined spatial–contextual information (ESA + SMLR) characterizes a state-of-the-art contribution in the research field of material identification. The proposed approach is exposed to present proper classification of the minerals of Mars surface in both the spatial and the spectral domain in short span of time.

Keywords Evolutionary multifactorial spectral analysis · Hyperspectral image classification · Mineral identification · Spectral analysis

N. Munusamy
Department of Computer Studies, KSG College of Arts and Science, Coimbatore, Tamil Nadu, India

R. P. Karchi (✉)
Bharathiar University, Coimbatore, Tamil Nadu, India

Abbreviations

SMLR	Sparse multinomial logistic regression
ESA	Evolutionary multifactorial spectral analysis
OMEGA	Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité
ENVI	Environment for Visualizing Images

81.1 Introduction

Hyperspectral images of the surface and subsurfaces of Mars at different ranges vary in their spatial resolution. For 352 contiguous spectral channels between 0.3 and 4 km/pixel where these vary for the visible and near infrared ranges. For the near infrared spectral band on 256 frequency channels is between 0.3 and 5.2 mm and for visible range it is 128 channels. This information is useful to identify chemical species on the Mars surface and the Mars atmosphere more precisely than previously. Further, this new dataset provides an open investigation to find the mineral composition of Mars surface. It is done by giving right to use the attributes of mineral species. Several unsupervised classification techniques can be applied to analyse hyperspectral data using different constraints. The proposed work investigates the utility of a partially observed classification model in realizing new minerals on the Martian surface with the Omega images, as automated methods and analytical algorithms differ in their nature of methodology, where automated algorithms are used to detect interested area of dataset and also produce global maps. Further analytical algorithms use the information related to mineral absorption bands. The information with respect to spectral variations at the surface are provided by band ratios. Tetracorder algorithm [1] provides mineral classifications by customizing the spectral features. To categorize and measure mixtures of mafic minerals at the Mars surface Tetracorder algorithm has been adapted [2]. The algorithm helps to have composition maps of mafic minerals present on Mars surface. Principal component analysis and independent component analysis are considered as better evaluating methods to check for similarities between each pixel spectrum, and these pixels or endmembers are rich in sulphate [3]. Hence, these techniques have been demonstrated to check for spectrum mixture appropriate for the arrangement of the individual reflectance spectra of endmembers. The main assignment is to categorize the components that are set in spatially individual patterns, type of sand grains in areas that exhibit different mineralogies like syrtis, iron oxides, gypsum and ice. The proposed work presents a novel unsupervised segmentation algorithm named as evolutionary component analysis for remotely sensed hyperspectral image data for material identification in the spectral [4] and spatial information. Here, the algorithm SMLR is preferred to acquire the knowledge about posterior probability distributions from both spatial and spectral information of the images where an image contains class imbalance information. To infer such imbalance, class distribution of the

testing hyperspectral data is done with evolutionary multifactorial spectral analysis. This approach helps to characterize noise and highly mixed pixels with less training set with high training quality and utility with respect to spectral signatures and its spectral changes with less interaction for classification in a better way. Also it derives most extreme spectra in the image end-members. The proposed segmentation approach based on evolutionary multifactorial spectral analysis (ESA) is investigated and estimated using both real and simulated hyperspectral datasets. ESA is evaluated and compared with newly introduced classification methods for the endmember extraction in the mixed pixel by revealing up-to-date performance. The integration of ESA methods with the SMLR algorithm further combined spatial–contextual information that represents an inventive idea of work in the research field of material identification. The rest of the paper is organized as follows: Sect. 81.2 gives details about the background knowledge regarding the related work. Section 81.3 gives explanations and formulates the proposed system. The experimental end results are conferred in Sect. 81.4. Finally the conclusion and future work is presented in Sect. 81.5.

81.2 Related Work

The state-of-the art methods reported in the literature from several authors for hyperspectral image segmentation are summarized in the following in three categories.

Sparse Multinomial Logistic Regression (SMLR) Model The proposed work uses joint probability distribution with SMLR model. This helps to learn the class posterior probability distributions. By using multinomial logistic regression, it can be classified to identify the unknown components from the known components. Further it helps to generate the feature selection and detection.

Artificial Neural Networks The artificial neural network has chosen in the hyperspectral image classification for their facility to model complex shapes with numerical weight and without analytical description. It contains set of adaptive weights and it is capable of approximating the spectral signatures. The following is the output function of neural network $N_i = f[\sum w_j, p_j]$, where N_i is the output of the i^{th} unit, $f(\cdot)$ is the activation function where it bounds the resulting output values as either true or false by $[0,1]$ or $[-1,1]$, w_j indicates the weights of input and p_j indicates the j^{th} input. Here each perception will be sensitive to various features like colour, edges, etc. It works better with respect to the fast computing time.

Spectral Angle Mapping Spectral angle mapping is formed by the pixel values and each endmember vector. It helps to calculate the angle between vectors. The minimal spectral angle indicates the more similar a pixel belongs to endmember class. Spectral angle mapping is combined with the Pixel Purity Index with the aim of deriving the most extreme end-members in an image. These outcomes can further be

used as inputs to extract the features for the classifications. It can also be used to detect the spectral changes in the images [5–10]. Spectral angle mapping can be done with following steps: (a) load the data image/dataset, (b) read sample end-member spectra using dataset, (c) calculate spectral angle formed by the pixel values and each end-member vector and (d) classify the end-members based on minimal spectral angle

81.3 Proposed Model of Evolutionary Multifactorial Spectral Analysis (ESA)

The ESA is deliberated to compare an unknown spectrum using multifactorial conditions and constraints. Each pixel of an image is processed separately and it is treated as individual. The evolutionary algorithm is based on spectrum population generation, where ‘X’ and ‘Y’ are a vector containing the coefficients of each component of the input library and vector having a spectrum, where ‘A’ is a ‘N*M’ (i.e. N-number of endmembers, M-number of spectral channels) matrix filled with the end-member spectra which can be considered candidate pool with possible number of the candidate solutions, ‘V’ is the vector of remote spectral region, individual A is the matrix containing the candidate pool, candidate pool is the number of end-members containing the candidate solutions and candidate solution is a spectral channel where fitness value is the maximum number of spectral end-members. Fitness value eliminates the noise features. Fitness value decides the quality of data classification. Figure 81.1 describes the different phases of the proposed evolutionary component analysis method.

Unmatched pixel is considered as residuals. Here the uncertainty inborn from physical measurement is compared with residuals. Whereas the uncertainty factors such as inherent level of noise and the errors of residual calibration (the non-linearity of the OMEGA detector) are included in image with respect to the “Observatoire pour la Minéralogie, l’Eau, les Glaces et l’Activité” (OMEGA) spectrometer. There exists uncertainty of 0.05 level of significance traditional estimation of the band detection threshold with OMEGA data. Further the Planetary Science Archive OMEGA website provides the information as the calibration documents.

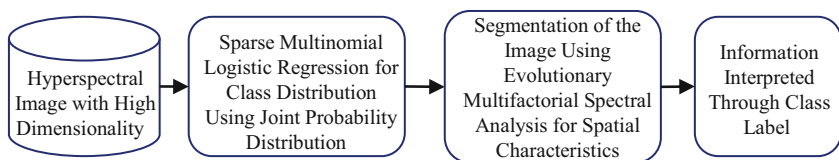


Fig. 81.1 Process flow of the ESA technique

81.3.1 A Subsection Sample Selection of Spectral Features: Mineral Selection for Classification

The spectral features tend to be minerals with measured surface spectrum in the MARS image. The input library can be built from the known composition or from the dataset itself.

The main criterion to detect the mineral compositions in the spectral field of interest depends on the existence of absorption bands. The investigators explained about spectral domain minerals which have very well varying endmember or even no specific spectral features where it is often more than the spatial variability of geological units [11, 12]. Mafic minerals (olivine) contain phyllosilicate end-members. Further they lead to possible confusion between several mineral compositions of respective area. To overcome such confusion, the proposed approach works in identifying such minerals. According to the definition of linear regression, the standard alternate for identification of mineral end-members is identifying an unknown spectrum against a library of known spectra. A number of spectral libraries have been compiled which can be used for this purpose.

81.3.2 Classification of the Hyperspectral Features

To classify the Spectral Signatures, spectral coefficients will help through matching absorption based on the band similarity between the remote spectral feature and reference feature. Classification may fail due to variations in band depth as well as mineral abundance. The position spectral features are described using Eq. (81.1).

$$\begin{bmatrix} DN_1' \\ DN_2' \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} DN_1 \\ DN_2 \end{bmatrix} \quad (81.1)$$

where DN1 and DN2 are the positions of the spectral feature 1 and feature 2.

However, crossover coefficients can modify the average level of a spectrum in various ways by adding weight to pixel. Certainly, minor grain sizes imply a lesser crossover coefficient while scattering necessitates a contribution for a higher global shape. The position of the spectral feature 1(DN1) and the position of the spectral feature 2 (DN2) computed by using Eq. (81.1) are used to estimate the variance by employing Eq. (81.2) and covariance features using Eq. (81.3).

$$\text{Variance } v_{i,i} = \sum_{k=1}^n (DN_{i,k} - D\bar{N}_i)^2 / n - 1 \quad (81.2)$$

$$\text{Co variance } \nu_{i,j} = \sum_{k=1}^n (DN_i - D\bar{N}_i)(DN_j - D\bar{N}_j)/n - 1 \quad (81.3)$$

A flat spectrum with values ($\neq 0$) is essential to compare the deviations of the global shape. These values are also interpretations for pure variations due to shading and shadowing effects or featureless dark or bright minerals. It can be considered as evolutionary component, can be taken for iteration and will be classified into any class of maximum probability. The algorithm 1 describes the methodology of the evolutionary component analysis.

Algorithm: ESA

1. Begin
2. $P = [G/2]$ // G is a Groups
3. Choose representatives of groups randomly (GR)
4. $S = P$ // generate empty set
5. Randomly generate m strings with n length
6. for $i = 1$ to P do
7. Perform task T_j (using factorial rank r_{ij} of Pixel)
8. $\varphi_i = 1/\min\{r_{i1}, r_{i2}, \dots, r_{ik}\}$ (fitness function)
9. $\tau_i = \text{argmin}_j\{r_{ij}\}$.(skill factor) else
10. $i = i + 1$ then repeat step 8 and step 9
11. Find coefficient values (variance, covariance as cross over function)
12. Choose the best value and classify it w.r.t class
13. End

Here for every individual pixel in a population P, it is necessary to define factorial rank r_{ij} on task T_j . Further, define scalar fitness φ_i of pixel which is based on its best rank over all tasks ($\varphi_i = 1/\min\{r_{i1} \text{ to } r_{ik}\}$). Skill factor τ_i of pixel is the best task among all ($\tau_i = \text{argmin}_j\{r_{ij}\}$). It is necessary to quantify and associate the fitness of population values to design a multifactorial evolutionary algorithm in a multitasking environment [13, 14].

81.4 Experimental Analysis

The experiments are carried out with the environment for visualizing images (ENVI) tool. An image `hsv0001f9c0_01_ra209s_trr3` Mars dataset captured from OMEGA instrument is used for experimentation. The class probability distribution is well-read by using SMLR to produce the selection of features to identify the subset of the smaller groups of the identical features. Hence feature extraction is done with SMLR. Proposed work uses ESA for classification only, as it is time consuming for feature extraction. The input image is considered with a band-123 as shown in Fig. 81.2a.

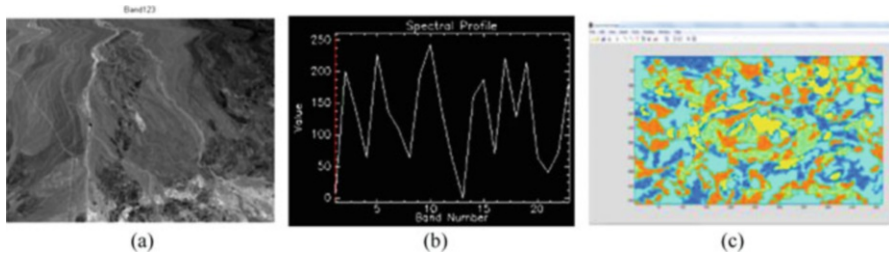


Fig. 81.2 (a) Input image (hsv0001f9c0_01_ra209s_trr3). (b) Spectral information of input image. (c) Classification of mineral compositions of Mars data

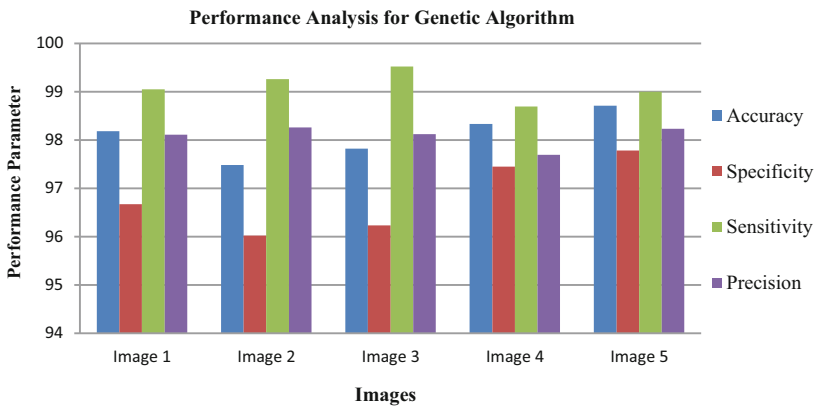


Fig. 81.3 Performance analysis graph for genetic algorithm

The input image bears its respective wavelength [15, 16] and band number which is shown in Fig. 81.2b. Further the input image is processed by SMLR to extract the features of end-members using multiple logistic regression. Then to have better classification, the proposed work introduced an ESA algorithm to select and classify the mineral composition present on the Mars surface which is shown in Fig. 81.2c.

The hyperspectral images are considered with constant spatial dimensions. Initially the image is pre-processed for the removal of noise and further processed for feature extraction. Hyperspectral information of the image is derived through SMLR for the generation of features. Further ESA algorithm is processed by using fitness function. To differentiate the classes by mean and variance criteria, it is necessary to go for selection and extraction of features. Further, Fig. 81.3 shows the performance of genetic algorithm in terms of accuracy, precision, recall and specificity. In the future the symbolic similarity analysis [17] can be employed for identification of the different mineral composition on Mars surface.

81.5 Conclusion and Future Scope

Hyperspectral image segmentation using evolutionary multifactorial spectral analysis for OMEGA dataset is achieved by an evolutionary multi-factorial method that includes two major steps. Initially, the proposed work gives emphasis to the class posterior probability distributions by adopting SMLR model with joint probability distribution. Further, the method of obtaining data in the initial step to segment the hyperspectral image using an evolutionary technique. By using spatial information it is compatible to estimate the dependencies between the endmembers. The enactment of the proposed method is explored by using the evolutionary multifactorial hybrid approach (i.e. SMLR and ESA). Henceforth, the experimentation is to be performed with Mars imagery datasets collected by using other instruments.

References

1. Clark RN, Swayze GA, Wise R, Livo KE, Kokaly RF, Sutley SJ, Dalton JB, McDougal RR, Gent CA (2003) Imaging spectroscopy: earth and planetary remote sensing with the USGS Tetracorder and expert systems. *J Geophys Res* 108(E12):5–1
2. Dalton JB, Sutter B, Kramer MG, Stockstill KR, Moersch J, Moore J (2003) Searching for aqueous mineralogy on Mars utilizing a surface water flow model Mars global surveyor and Mars odyssey data. American Geophysical Union, Fall Meeting, Abstract #P21B-0052
3. Ezhilarasi, M., Krishnaveni V. “An evolutionary multipath energy-efficient routing protocol (EMEER) for network lifetime enhancement in wireless sensor networks” *Soft Computing*, (2019). <https://doi.org/10.1007/s00500-019-03928-1>
4. Forni O, Poulet F, Bibring JP, Erard S Gomez C, Langevin Y, Gondet B (2005) The Omega Science Team, 2005. Component separation of OMEGA spectra with ICA. In: 36th annual lunar and planetary science conference, March, abstract no.1623,
5. Angadi SA, Hatture SM (2011) A novel spectral graph theoretic approach to user identification using hand geometry. *Int J Mac Intell* 3(4):282–288
6. Karchi RP, Munusamy N (2018) Hyperspectral image classification and Unmixing by using ART and SUnSPI techniques. *Int J Database Theory Appl* 11(3):13–28
7. Moussaoui S, Hauksdóttir H, Schmidt F, Jutten C, Chanussot J, Brie D, Douté S, Benediktsson JA (2008) On the decomposition of Mars hyperspectral data by ICA and Bayesian positive source separation. *Neurocomputing* 71:2194
8. Shuyuan Y, Qiao Y, Yang L, Jin P, Jiao L (2014) Hyperspectral image classification based on relaxed clustering assumption and spatial Laplace Regularizer. *IEEE Geosci Remote Sens Lett*
9. Combe J-P, Le Mouélic S, Sotin C, Gendrin A (2008) Analysis of OMEGA/MarsExpress data hyperspectral data using a multiple-endmember linear spectral Unmixing model (MELSUM): methodology and first results. *Planet Space Sci* 56:951
10. Karchi RP, Munusamy N (2018) Exploration of Unmixing and classification of hyperspectral imagery. *Int J Futur Gener Commun Netw (IJFGCN)* 11(6):13–31
11. Angadi SA, Hatture SM. [Hand Geometry Based User Identification Using Minimal Edge Connected Hand Image Graph](#)(2018) *IET Comput Vis* 12(5):744–752
12. Bibring J-P, Soufflot A, Berthe M, Angevin Y, Gondet B, Drossart P, Bouye M, Combes M, Puget P, Semery A, Bellucci G, Formisano V, Moroz V, Kottsov V (2004) The OMEGA Co-I team, 2004. *Eur Space Agency Spec Pb* 1240:37

13. Bibring J-P, Langevin Y, Gendrin A, Gondet B, Poulet F, Berthe M, Soufflot A, Arvidson R, Mangold N, Mustard JF, Drossart P (2005) The OMEGA team, 2005. Mars surface diversity as revealed by the OMEGA/Mars express observations. *Science* 307(5715):1576–1581
14. Gu YF, Feng K (2012) L1-graph Semisupervised learning for hyperspectral image classification. In: *IEEE proceedings of international geosciences remote sensing symposium*, Munich, Germany, pp 1401–1404,
15. Gupta A, Ong YS, Feng L (2015) Multifactorial evolution: towards evolutionary multitasking. *IEEE Trans Evol Comput* 20(3):343–357
16. Nagarajan M, Karthikeyan S (2012) A new approach to increase the life time and efficiency of wireless sensor network. In: *Proceedings of IEEE international conference on pattern recognition, informatics and medical engineering*
17. Ezhilarasi M, Krishnaveni V (2018) A survey on wireless sensor network: energy and lifetime perspective. *Taga J* 14:3099–3113
18. Angadi SA, Hatture SM (2019) Face recognition through symbolic modelling of face graphs and texture. *Int J Pattern Recognit Artif Intell.* <https://doi.org/10.1142/S0218001419560081>

Chapter 82

Novel Lifting Filter Bank for Bird Call Analysis



N. Subbulakshmi and R. Manimegalai

Abstract Lifting scheme is widely used in signal processing applications due to advantages such as fast implementation, less computational complexity, and perfect reconstruction (Cohen A, Daubechies, I, Feauveau J, Commun Pure Appl Math, 1992). The proposed work discusses the design of a Novel Lifting-based Filter Bank (NLFB) for analyzing bird calls and songs. The proposed filter bank has four lifting steps such as split, predict, update, and merge that are used for perfect decomposition and reconstruction. The proposed NLFB consumes 65% less area and 17% less power when compared to interpolated filter bank.

Keywords Signal processing · Filter bank · Lifting scheme · Bird call analysis

Abbreviations

CDF	COHEN-DAUBACHES-FEAUVEAU
dB	Decibels
MVFB	Modified Variable Filter Bank
NLFB	Novel Lifting-based Filter Bank
SNR	Signal to Noise Ratio

N. Subbulakshmi (✉)
Malla Reddy Engineering College (A), Hyderabad, Telangana, India

R. Manimegalai
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_82

82.1 Classification of Lifting Techniques

A typical classification of lifting techniques is shown in Fig. 82.1. They are classical lifting, space-varying lifting, multidimensional lifting [1], integer lifting [2], adaptive lifting [3], and vector lifting. Filter banks and filters are regularly designed as one-dimensional and are converted into two-dimensional by filtering. Space-varying lifting selects the filter at each sample n according to local signal characteristics. Lifting filters are designed to compute the multidimensional wavelet transform. Multidimensional lifting is used in order to increase the number of vanishing moments of the wavelets and retain the signal mean on every level. In most of the designs, filters include floating point coefficients. If the input data has sequence of integers, then the resulting filter output will not have integers. This problem has been resolved using Fourier and cosine transform in [4]. Adaptive lifting is the modified version of the classical lifting. Vector lifting provides two compact multi-resolution representations in image coding.

Perfect reconstruction of filter bank and Euclidean algorithm make lifting possible. Lifting has advantages over Fourier transform since it works on both time domain and frequency domain [11]. Data is divided into two smaller parts, predict the parts in the original data and update the detail as the difference between data and prediction. Lifting version of CDF (2, 2) is also known as (5, 3) integer lifting. Lifting-based methods require less computation when compared to conventional approaches. In data compression, if size of the speech signal is less than 10 KB, then, the signal can be compressed without dividing into blocks. If the signal is too large, then, the signal is divided into sub-blocks.

82.2 Cohen-Daubach-Feauveau (CDF) Wavelet Filter

CDF wavelet is a category of Biorthogonal wavelet [5]. The scaling function of CDF is always symmetric. It can be implemented using lifting scheme and it is used for many multimedia applications such as image, video, and audio. Real time signal compression using lifting scheme is implemented by Truong Phong Tuyen 2015. Applications of lifting scheme are discussed in [6]. Haar, Cubic, CDF, and Daubechies-4 series such

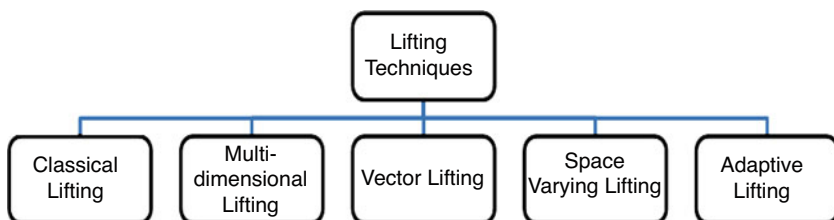


Fig. 82.1 Classification of various lifting techniques

as Daub-4, Daub-6, Daub-8, Daub-10, Daub-12, and Daub-16 are some of the available lifting schemes. Biorthogonal CDF consists of the following types: (i) CDF (5, 3), (ii) CDF (7, 5), and (iii) CDF (9, 7). Table 82.1 shows the number of coefficients in analysis and synthesis bank. In CDF (5, 3), the number of analysis and synthesis filter bank coefficients is 5 and 3, respectively. The decomposition process using CDF technique is shown in Figure 82.2. Data samples are divided into even and odd samples. Then, they are processed using predict and update steps. First, the predict sequences are obtained as new odd samples. The newly generated odd samples are taken as the input for the update process. Figure 82.3 shows the reconstruction of the signal. The approximations and details are given to the update and predict blocks. In the update process, the new even samples are produced. Those even samples are used in prediction process to generate new odd samples.

82.2.1 Filters Used for Bird Call Analysis

In late decades, a few clamor evacuation and sound upgrade strategies have been proposed for handling human discourse signals. As of late, there has been expanding interest in approaches to naturally perceive flying creature species in natural

Table 82.1 Coefficients in the design of CDF

Type	Number of analysis coefficients	Number of synthesis coefficients
CDF (5,3)	5	3
CDF (7,5)	7	5
CDF (9,7)	9	7

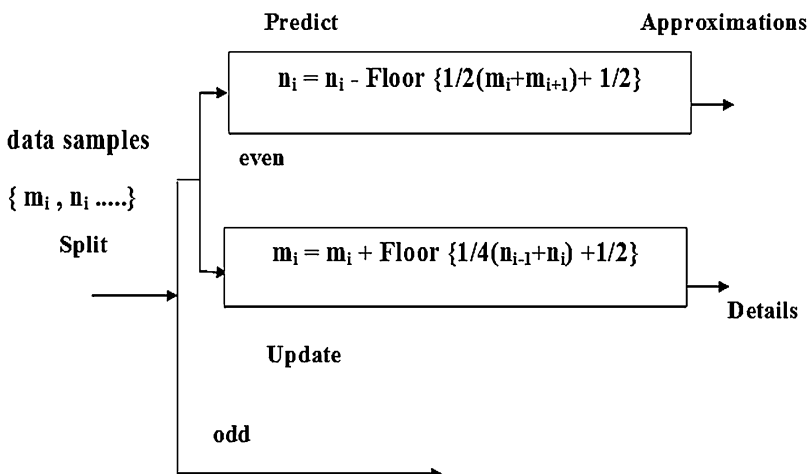


Fig. 82.2 CDF decomposition

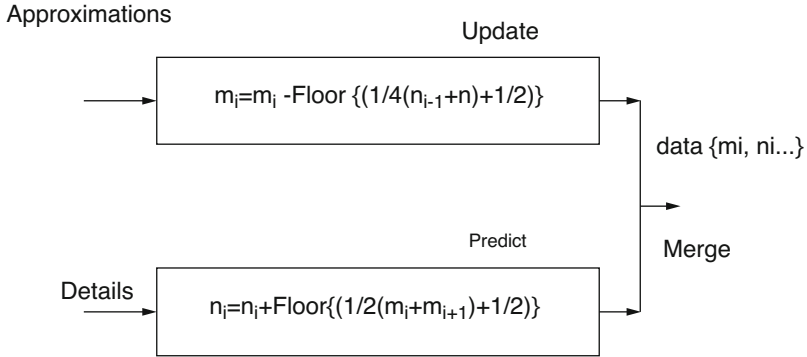


Fig. 82.3 CDF reconstruction [11]

chronicles. Clamor obstruction has been a significant issue in this exploration territory as it can conceivably diminish the exactness of flying creature acknowledgment. The least difficult ways to deal with diminishing foundation clamor in sound chronicles are low and high-pass filtering. These filters lessen frequencies in locales of sound known not to contain any flag. With regard to bioacoustics, the calls of creatures of intrigue are regularly known to be in a certain recurrence extend, so anything not in this recurrence range can be wiped out. Feathered creatures commonly do not make sounds over 12 kHz or beneath 1 kHz [17], so sounds outside of this area can be overlooked. Neal et al. [12] utilize a 1 kHz high-pass filter as a major aspect of a push to portion winged animal sounds. Cook and Logue [13] utilize indistinguishable methodology from part of a procedure to think about contrasts between ongoing decades; a few clamor evacuation and sound upgrade techniques have been proposed for preparing human discourse signals. Of late, there has been expanding interest in approaches to naturally perceive winged creature species in natural chronicles.

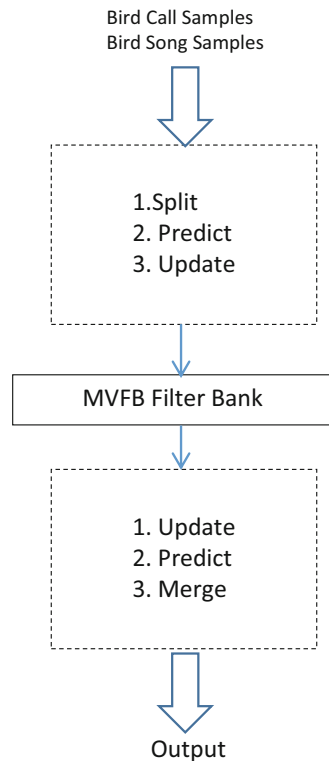
Commotion obstruction has been a significant issue in this examination territory as it can conceivably diminish the precision of winged creature acknowledgment. The most straightforward ways to deal with lessening foundation clamor in sound accounts are low and high-pass filtering. These filters constrict frequencies in areas of sound known not to contain any flag. With regard to bioacoustics, the calls of creatures of intrigue are regularly known to be in a certain recurrence run, so anything not in this recurrence range can be dispensed with. Flying creatures regularly do not make sounds over 12 kHz or beneath 1 kHz [17], so sounds outside of this locale can be overlooked. Neal et al. [12] utilize a 1 kHz high-pass filter as a component of a push to portion flying creature sounds. Pastry specialist and Logue [13] utilize indistinguishable methodology from part of a strategy to think about contrasts between chickadee sounds crosswise over populaces. Additionally, Bardeli et al. [14] utilize low-pass filtering to help identify two jeopardized winged creature species. Notwithstanding, since they forcefully expel any solid from target recurrence areas, they can be utilized in blend with other procedures to enhance clamor decrease [7].

82.2.2 Lifting Scheme Algorithm

Lifting scheme algorithm is basically used to generate second-generation wavelet in image compression. It is suitable for noise reduction. The signal decomposition and reconstruction are made using the lifting steps, namely, splitting, prediction (P), and update (U) [5]. Perfect reconstruction filter bank and the Euclidean algorithm play important roles in lifting algorithm [5]. However, Fourier analysis is not used in lifting. First, the high-pass is converted into low-pass filters using polyphase matrix. Then, the Euclidean algorithm is applied to decompose the matrices. The matrix elements are used as coefficients to develop prediction and update. For the signal transformation, the input data samples are decomposed into high- and low-frequency components. The low-frequency components are denoted as approximations. The high-frequency components are termed as details. Figure 82.4 shows the lifting steps used for the analysis filter bank. Equations (1) and (2) are used to predict and update in CDF method [5].

$$P(M_i) = [1/2(M_i + M_{i+1}) + 1/2] \quad (82.1)$$

Fig. 82.4 Proposed NLFB for bird analysis



$$U(N_i) = [1/4(N_i - 1 + N_i) + 1/2] \quad (82.2)$$

$M_i + 1 = X_{\text{even}} = \text{even data from input samples}$

$N_i + 1 = X_{\text{odd}} = \text{odd data from input samples}$

The multilevel decomposing and reconstruction trees are shown in Fig. 82.4. At every stage of decomposition, all sequences of data are split into odd and even sequences. Equation (82.3) states that after lifting, predict X_{odd} with X_{even} and replace X_{odd} by prediction error in multilevel signal decomposition. In Eq. (82.4), $M_i + 1$ is updated with $N_i + 1$ and replaced by updated factor. The above steps are repeated till all input sequences are obtained at the output.

$$N_i + 1 = N_i + 1 - P(M_i + 1) \quad (82.3)$$

$$M_i + 1 = M_i + 1 + U(N_i + 1) \quad (82.4)$$

82.2.2.1 Algorithm for First-Level Decomposition (Fig. 82.4)

a. Splitting: Even $M_1 = \{2, 4, 6, 8\}$

Odd $N_1 = \{1, 3, 5, 7\}$

b. Predict:

$$\begin{aligned} n_1 &= n_1 - \lfloor (1/2 (m_1+m_2) + 1/2) \rfloor \\ &= 1 - \lfloor (1/2 (2+4) + 1/2) \rfloor \\ &= -2 \end{aligned}$$

Similarly,

$$\begin{aligned} n_2 &= n_2 - \lfloor (1/2 (m_2+m_3) + 1/2) \rfloor = -2 \\ n_3 &= n_3 - \lfloor (1/2 (m_3+m_4) + 1/2) \rfloor = -2 \\ n_4 &= n_4 - \lfloor (1/2 (m_4+m_5) + 1/2) \rfloor = 3 \\ N_1 &= \{-2, -2, -2, 3\} \end{aligned}$$

c. Update:

$$\begin{aligned} m_1 &= m_1 + \lfloor (1/4 (n_0+n_1) + 1/2) \rfloor \\ &= 2 + \lfloor (1/4 (0+(-2)) + 1/2) \rfloor \\ &= 2 \end{aligned}$$

Similarly,

$$\begin{aligned} m_2 &= m_2 + \lfloor (1/4 (n_1+n_2) + 1/2) \rfloor = 3 \\ m_3 &= m_3 + \lfloor (1/4 (n_2+n_3) + 1/2) \rfloor = 5 \\ m_4 &= m_4 + \lfloor (1/4 (n_3+n_4) + 1/2) \rfloor = 8 \\ M_1 &= \{2, 3, 5, 8\} \end{aligned}$$

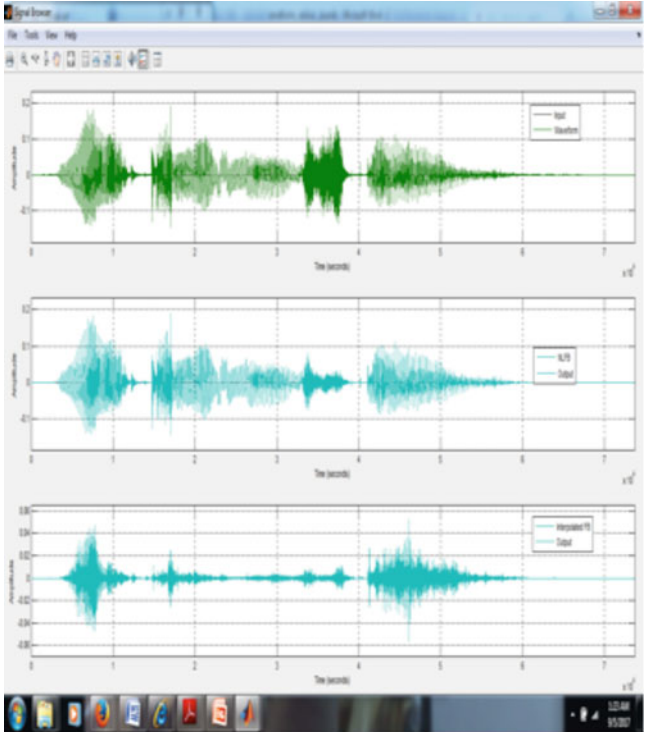
82.2.3 Bird Call Analysis

For the present investigation, a method utilized a subset of on-fowl sound chronicles, got amid an alternate report (Gill et al., in planning) [4]. The information were gathered in the South of Germany, from 12 singular jackdaws (*Corvus monedula*, 7 hostage housed and 6 free-living), right off the bat in the long stretches in 2014 and 2015. Knapsack application was endorsed by the Government of Upper Bavaria and in consistence with the European mandates for the insurance of creatures utilized for logical purposes (2010/63/EU). The rucksacks comprised of a monetarily accessible advanced voice recorder (Edic Mini Tiny A31, TS-Market Ltd., Russia), a battery-powered battery (ICP581323PA to ICP402035, Renata, Switzerland), a radio transmitter for movement (BD-2 Holohil, Canada), and a contracting tube packaging. The developers were charged, modified, and read out by means of PC connection and concurring programming (RecManager, rendition 2.11.19, TeleSystems, Russia). They were set to record consistently for a couple of hours each morning, for a couple of days, starting one day post catch (at 22050 Hz examining rate, uncompressed.wav organize). This gave cognizant vocalization information and acoustic foundation data, rather than utilizing sufficiency-based triggers (however at an expense of capacity and battery).

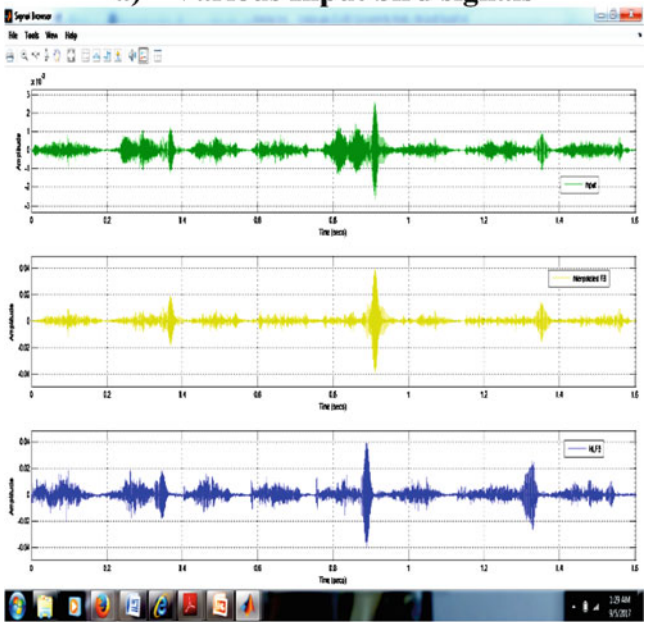
For knapsack connection, feathered creatures were either prepared to fly inside a littler compartment of the aviary where they were found utilizing winged creature nets (imprisonment) or caught inside their home boxes (wild). Knapsacks were fitted utilizing endorsed connection strategies [8] and following normal suggestions (<5% of body weight; near focus of gravity). Flying creatures were independently recognized by shading rings. After catch and knapsack connection (20 mins \pm 4.1 SD), they were watched utilizing binoculars and additionally radio-telemetry, what not of them were instantly ready to fly upon discharge. For further points of interest on strategies and creature welfare, see Gill et al. (in planning) [7, 9, 10] (Fig. 82.5).

In order to determine the odd and even signals, two half-band filters, i.e., low-pass and high-pass, are used. In third level decomposition, one odd sample and one even sample are received after the prediction and update steps. In the synthesis bank, the reverse process is carried out through the update, predict, and merge steps. Table 82.2 shows the input signal and lifting coefficient. Stop band attenuation deteriorates because of coefficient decimation technique.

Comparisons based on various parameters such as area, delay, and power for the interpolated filter bank using the proposed NLFB are shown in Table 82.2. The MVFB filter bank is used in the implementation of the proposed filter bank design and tested using various bird calls. All kinds of inputs such as bass, low midrange, midrange, and upper midrange frequencies are given as the samples to verify the functionality of the proposed filter design. Signal-to-noise ratio is also an important part in signal processing. Table 82.3 shows the SNR measurement for the proposed filter bank (Fig. 82.6).



a) Various Input bird signals



b) Corresponding Output Signals

Fig. 82.5 Input and output of the proposed filter bank (a) Various input bird signals,, (b) Corresponding output signals

Table 82.2 Input signal and lifting coefficient

Input signal	Filter coefficient
Sample 1 $f_s = 11.025\text{KHz}$	$H1 = [-0.006 \ 0 \ 0.0109 \ 1.000 \ 0.995 \ 0 \ 0.00780 \ -0.0141 \ 0 \ 0.0072]$ $H2 = [0 \ 0 \ 0 \ -0.550 \ 1.000 \ -0.5076 \ 0]$
Number of samples = 2712	$F1 = [0 \ 0 \ 0 \ 0.550 \ 1.000 \ 1.000 \ 0.5076]$ $F2 = [0 \ 0 \ 0 \ -0.0109 \ 0.9945 \ 0 \ -0.9859 \ 0.0072]$
Sample 2 $f_s = 11.025\text{KHz}$	$H1 = [-0.456 \ 0 \ 0.0966 \ 1.000 \ 0.9514 \ 0 \ 0.0271 \ -0.0574 \ 0 \ 0.0289]$ $H2 = [0 \ 0 \ 0 \ -0.4722 \ 1.0000 \ -0.5033 \ 0]$
Number of samples = 9000	$F1 = [0 \ 0 \ 0 \ 0.4722 \ 1.0000 \ 1.0000 \ 0.5033]$ $F2 = [0 \ 0 \ 0 \ -0.0966 \ 0.9514 \ 0 \ -0.9426 \ 0.0289]$

Table 82.3 Comparison based on SNR

Signal-to-noise ratio for varying noise levels									
Filter/input noise bank (in dB)	5	10	15	20	25	30	35	40	45
Proposed NLFB	9.12	12.24	14.36	19.13	22.98	24.81	25.11	26.56	28.77
Interpolated filter bank	8.02	11.30	12.32	14.21	20.05	23.11	24.73	25.23	26.34
ANSI filter bank	6.38	10.54	11.42	13.87	18.66	22.34	24.68	26.54	27.81
Reconfigurable filter bank	9.23	12.65	13.24	19.77	18.32	23.13	25.17	27.86	29.05

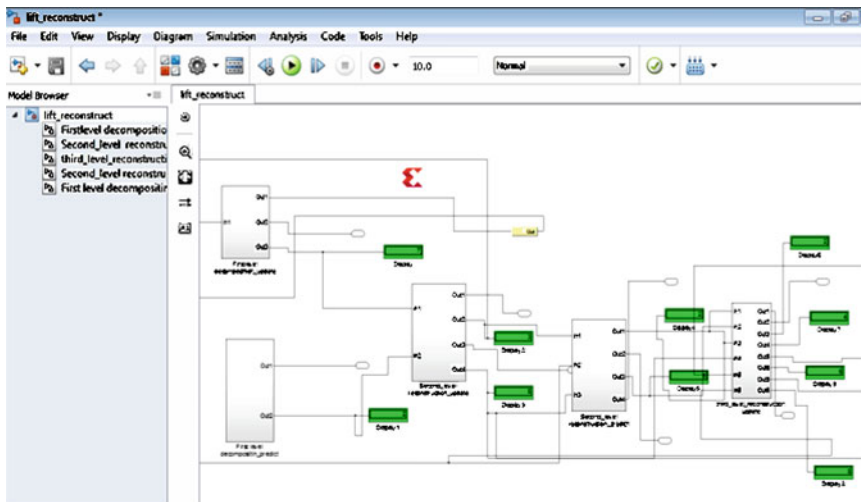


Fig. 82.6 Second level reconstruction using the proposed NLFB

82.3 Conclusion

The proposed novel lifting-based filter bank, NLFB, for analyzing bird sounds includes four lifting steps such as split, predict, update, and merge. It has decomposition and reconstruction steps at the analysis bank and synthesis bank, respectively. The core filter design using the proposed MVFB enhances the performance of the design when compared to the interpolated filter bank. The proposed filter bank provides perfect reconstruction of the applied signal at the output. The matrix elements are used as coefficient to develop prediction and update. The perfect reconstruction of the applied signal is achieved at the output using the proposed filter bank design. The proposed NLFB consumes 45 mW power and has 91.142 ns delay when compared to interpolated filter bank design.

References

1. Ding W, Xiong R, Shi Y, Kong D, Yin B (2011) Fast mode dependent directional transform via butterfly-style transform and integer lifting steps. *J Vis Commun Image Represent* 22 (8):721–726
2. Cristina B, Cioffi F, Guida M, Roggero M (2016) The scheme of liftings and applications. *J Pure Appl Algebra* 220(1):34–54
3. Verdicchio F, Andreopoulos Y (2011) Distortion estimates for adaptive lifting transforms with Nois. *Image Vis Comput* 29(11):744–758
4. Hong J (1993) Discrete fourier, hartley and cosine transforms in signal processing, Ph.D. thesis, Department of Electrical Engineering, Columbia University
5. Cohen A, Daubechies I, Feauveau J (1992) Biorthogonal bases of compactly supported wavelets. *Commun Pure Appl Math* 45(5):485–560
6. Cheng-yi X, Hou J-h, Tian J-w, Liu J (2007) Efficient Array architectures for multi-dimensional lifting-based discrete wavelet transforms. *Signal Process* 87(5):1089–1099
7. Jeon KM, Lee DY, Kim HK, Lee MJ (2014) Acoustic surveillance of hazardous situations using nonnegative matrix factorization and hidden Markov model. In: *Audio engineering society convention* 137. Audio engineering society
8. Cobos M, Perez-Solano JJ, Felici-Castell S, Segura J, Navarro JM (2014) Cumulative-sum-based localization of sound events in low-cost wireless acoustic sensor networks. In: *IEEE/ACM transactions on audio, speech, and language processing*, vol 22(12), pp 1792–1802
9. Kucukbay SE, Sert M (2015) Audio-based event detection in office live environments using optimized mfcc-svm approach. In: *Semantic computing (ICSC), IEEE international conference on*, IEEE, pp 475–480
10. Dufaux A (2001) Detection and recognition of impulsive sounds signals. *Institute de Micro-technique Neuchatel, Neuchâtel*
11. Subbulakshmi N, Manimegalai R (2017) NLFB: design of a novel lifting based filter bank for digital hearing aid. *J Comput Theor Nanosci* 14(3):1410–1416

Chapter 83

Automatic Classification of Solid Waste Using Deep Learning



V. P. Brintha, R. Rekha, J. Nandhini, N. Sreekaarthick, B. Ishwaryaa, and R. Rahul

Abstract Solid waste management is an essential task to be carried out in day-to-day life. So an automated recognition system using deep learning algorithm has been implemented to classify wastes as biodegradable and non-biodegradable. Efficient segregation of solid wastes helps to reduce the amount of waste buried in the ground, thereby improving the recycling rate, and safeguards the soil from pollution.

Keywords Waste · Glass · Plastic · Wood · Textile · Classification · Accuracy

Abbreviations

CNN Convolutional Neural Network
R-CNN Region Convolutional Neural Network

83.1 Introduction

Solid waste management has turned out to be one of the fundamental issues in both urban and rustic regions. The significant increase in municipal solid waste generation has been recorded worldwide. Most urban solid waste in Indian cities and towns is land filled and dumped. Currently there is no system for segregation of dry, wet and metallic wastes. An efficacious management needs to be materialized for replacing traditional way of dealing with waste. Current worldwide waste generation levels are roughly 1.3 billion tonnes for each year and are relied upon to increase around 2.2 billion tonnes every year by 2025. India produces 1,00,000 metric huge amounts of waste every day. Strong waste administration was made a need in the

V. P. Brintha · R. Rekha (✉) · J. Nandhini · N. Sreekaarthick · B. Ishwaryaa · R. Rahul
Department of Information Technology, PSG College of Technology, Coimbatore, Tamil Nadu, India

National Mission on Sustainable Habitat. The current refuse transfer framework in India comprises of unclassified waste gathered from homes which are then isolated at a station physically.

The segregation of solid waste done by manual labour can bring about many health hazards for the waste sorters in addition to being less efficient, time consuming and not completely feasible due to their large amount. In this project, an automated recognition system using Deep learning algorithm has been implemented to classify wastes as biodegradable (Agricultural Wastes, Food Wastes, Clothes, Paper, Mineral Wastes) and non-biodegradable (Plastics, Steel, Metal Cans), where the system once trained with an initial dataset, can identify objects real-time and classify them almost accurately. Efficient segregation of solid wastes helps to reduce the amount of waste buried in the ground, thereby improving the recycling rate, and safeguards the soil from pollution. Biodegradable waste is utilized for power generation, soil enrichment and food for animals. This process does not harm the earth but instead makes it valuable, ecologically safe and helps to protect the environment, ecosystem and human inhabitants in future.

83.1.1 Objective

The main objective of this project is to develop various computer vision approaches to classify garbage in an efficient manner into recycling categories to process waste. The images of a single piece of recycled material or garbage are taken and classified into six classes consisting of glass, paper, metal, plastic, cardboard and trash. A dataset that contains around 400–500 images for each class, which was hand collected, has been created. The model used is faster convolutional neural network (CNN). The performance of CNN was good; however, the CNN was not trained to its full capability due to trouble finding optimal hyperparameters.

83.1.2 Scope of the Proposed Methodology

The trend to instil automated waste segregation systems into new housing development projects is rising and serves for leading the way in implementing this type of technology. Statistics show that the UK produces 330 million tonnes of waste a year, and estimates show that in 2007, 480,000 tonnes of CO₂ were emitted due to separate collection and composting of 835,000 tonnes of organic waste. This shows us the need for segregating and composting garbage on a larger scale. Hence this project helps inefficient segregation of solid waste and provides the right path for decomposition. The major scope of this project is to make our

environment sustainable, to maintain a safe and green environment and to reduce the human work in classifying the waste as degradable and non-biodegradable.

83.2 Related Work

The prevailing garbage disposal system in India consists of unorganized waste collected from habitats which are then segregated at a station. The segregation is done by manual labour which has many health mishaps for the labourers and is time consuming and also requires financial share to the workers. Another concept uses a hardware component that can sort waste at the initial stage thus making waste management more powerful. The designed system sorts wastes into six different categories, namely, metal, plastics, paper, textile, wood and glass. An OpenCV, which acts as the heart of the system, is then used to observe the wastes and timing of all the subsections so as to sort the waste into the three primary categories. Literature survey considering various papers is summarized in Table 83.1.

Table 83.1 Literature survey

Paper name	Problem solved	Demerits
An automatic classification method for environment: Friendly waste segregation using deep learning	A proposition for a system that automatically classifies waste using deep learning	Cannot classify medical waste and e-waste due to government restrictions
Comparing deep learning and support vector machines for autonomous waste sorting	To use SVM model to classify waste into plastic, paper and metal	Small amount of images is used in the training set
Automatic waste classification using computer vision as an application in Colombian high schools	Integrating computer vision into developing an application to classify waste automatically into the IEAB	The wrong classifications, the reduced waste and the images in the database indicate that this technique is not enough to carry out the classification in a system of the real scale
Municipal solid waste classification using microwave non-destructive testing technique	This paper presents basic researches of the variation of microwave signal propagation characterization to verify microwave is suitable for MSW classification	Resolution of microwave non-destructive detection of MSW is not always high enough and can be improved only by using stronger microwaves
Smart recycle bin: A conceptual approach of smart waste management with integrated web- based system	A smart recycle bin that caters for recycling glass, paper, aluminium cans and plastic products that automatically evaluate the value of the wastes thrown and accordingly provide 3R card	It is not a very energy-efficient process when the scale of the project is increased

83.3 Methodology

The proposed idea mainly concentrates on the classification and identification of the waste that is being dumped into the garbage. Usually, unsegregated waste is dumped in a landfill and they are made to decay which however takes thousands of years in the case of non-biodegradable waste. This project proposes an idea where a computer system on its own is able to identify the waste without any human intervention based on the material of the waste item, irrespective of its shape, colour and size, and classify them. The proposed system can learn by itself and hence can constantly update itself in case of new waste materials. The advantage of this proposed system is that it would include easy classification of solid wastes, decomposition of wastes, reduced health hazards and faster process that requires only a negligible initial investment for this to carry out. The block diagram of the proposed system is shown in Fig. 83.1.

83.3.1 Dataset Collection

Different types of wastes from the Internet are collected. To remove all the redundancies from those images, pre-processing will be done followed by training those datasets and finally developing a model and classifying the waste as biodegradable (textile, paper, glass, wood, waste, agricultural waste, food waste) and non-biodegradable (plastics, steel, metal cans).

The dataset collection process was done by capturing the waste images manually because there were no publicly available sites for datasets pertaining to waste materials. Initially the waste images collected from Google were used. However, these images do not accurately represent the kind of waste even after more research and progress on recycling wastes and the state of recycled waste goods. This is unlikely in recycled waste materials since they are treated as waste because they are unclean, improper, arbitrary shaped, crushed, etc. So the waste images were collected manually.

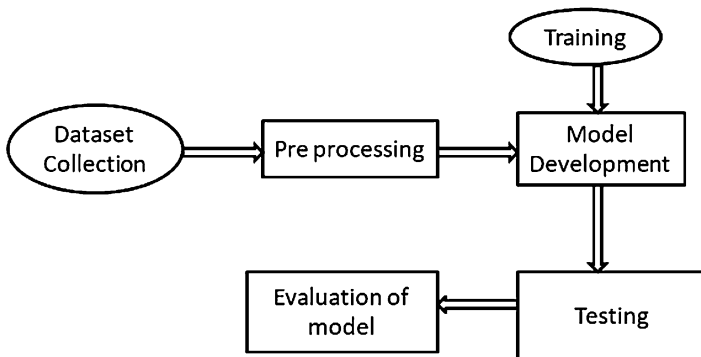


Fig. 83.1 Block diagram of the system

83.3.2 Processing Datasets

Few images of waste datasets are hand collected by us and the rest are taken from the Internet. The dataset contains images of solid waste objects across 7 classes with about 200–300 images in each type of waste, totalling about around 1100 images. The dataset collection process involved using a white paper as a background (even normal background is supported) and taking pictures of waste. The intensity of light and position for each image is not the same, which introduces variation in the dataset. Figure 83.2 shows example images of few classes of datasets. Various image transformation techniques were chosen to account for the different positions of recycled material which ultimately helps in increasing the quality of dataset.



Fig. 83.2 Types of wastes

83.3.3 *Types of Wastes*

The dataset contains seven classes that are mutually exclusive and the images are small, are clearly labelled and have no noise. The seven various types of wastes used are food waste, paper waste, plastic waste, metallic waste, wood waste, broken glasses, rubber and textiles. Some types of wastes are depicted in Fig. 83.2.

83.4 Implementation

83.4.1 *Algorithm Used: Convolution Neural Network*

A convolution neural network (ConvNet or CNN) is one of the most popular algorithms for [deep learning](#), a type of machine learning where a model learns to perform the task of classification directly from text, images, videos or sound. CNNs are specifically useful for finding the patterns in images to detect and recognize objects, faces and poses.

83.4.2 *Steps Involved*

Step 1: Start

Step 2: The system is powered on. The OpenCV comes up on the screen. The waste image is shown in front of it.

Step 3: Camera takes pictures of the objects to be identified. The interest points in the images are marked by a rectangle.

Step 4: Training data is used to analyse images. Only the marked points of test image are analysed with the training data.

Step 5: Detection and localization of objects by boundary algorithm that finds out the features in the images that could identify the image easily.

Step 6: Object identification takes place. The image is identified using the Fast R-CNN algorithm.

Step 7: Prediction: The prediction of type of waste is based on the identification and a probability index studied.

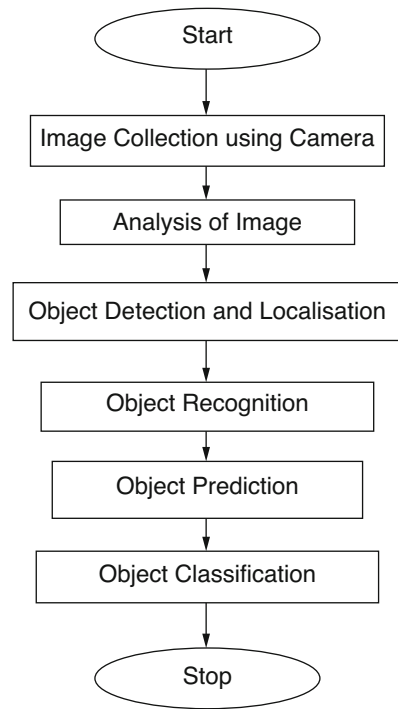
Step 8: Arriving at output: The prediction with highest probability is regarded as the output.

Step 9: Classification of object: The object is then classified to the type of waste it belongs to.

Step 10: Stop.

83.4.3 Flow Chart (Fig. 83.3)

Fig. 83.3 Flow chart



83.4.4 Model Development

Training is the first step to be practised after the collection of dataset. A large number of inputs have to be given to train the network. Here for the purpose of recognizing images, train a layer of features that receive input from the pixels of the image. The edges present in the image are recognized using Edge detection algorithm. Next, object parts are identified from various image processing algorithms such as feature extraction by faster R-CNN, Finally, objects are identified.

83.5 Experimental Results

The developed system could identify all the types of wastes both when shown individually as in Fig. 83.4 and also when they are mixed up together as in Fig. 83.5.

Fig. 83.4 Glass identification



Fig. 83.5 Mixed waste identification

Table 83.2 Result analysis

S.no.	Type of waste	Accuracy
1	Glass	99
2	Wood	67
3	Paper	97
4	Textile	70
5	Metal	70
6	Plastics	97

Table 83.2 Accuracy obtained on showing various types of wastes to the system.

83.6 Conclusion

The proposed system for the management of wastes is more efficient and time saving than the currently employed method that the municipality employees perform. Though this system is simple in concept, it is very valuable and affordable. Hence to ensure being automated, a system which takes a huge dataset with lots of images as input without human intervention and also has the capacity to think by itself offers the best solution. It acts as an aid for reducing pollution levels and in the long run focuses on the development of a nation and restoration of our ecosystem. Thus this project holds value as an important asset to the society.

References

1. Abdel-Hamid O, Mohamed AR, Jiang H, Penn G (2012) Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 4277–4280
2. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, vol 4, pp 1701–1708
3. Ciresan D, Meier U, Schmidhuber J (2013) Multi-column deep neural networks for image classification. In: Proceedings of the IEEE international conference on computer vision and pattern recognition (CVPR), IEEE conference, vol 8, pp 3642–3649
4. Smola A, Vishwanathan SVN (2008) Introduction to machine learning, vol 1. Cambridge University Press, Cambridge, pp 16–25
5. Chandramohan A, Mendonca J (2014) Automated waste segregator. In: Texas instruments India educators' conference (TIIEC). IEEE, Bangalore
6. Desai Y, Dalvi A (2009) Waste segregation using machine learning. Int J Res Appl Sci Eng Technol (IJRASET) 6:5
7. Vipin Upadhyay JAS, Poonia MP (2012) Solid waste collection and segregation. Int J Eng Innov Technol (IJEIT) 1:3
8. Aronszajn N (1950) Theory of reproducing kernels. Trans Am Math Soc 68(3):337–404
9. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst 2:1097–1105

Chapter 84

Relevancy and Similarity Aware Drug Comment Classification Framework on Social Media Drug Posts



D. Krithika Renuka and B. Rosiline Jeetha

Abstract Social media offer superior platform for its clients to share their comments and opinions about their knowledge toward specific product. This attracts the medical investigators to acquire knowledge about drug products owing to user perspective. Moreover, gaining and analyzing knowledge about drug tweet status is extremely a complex talk. This is carried out in our previous investigation by introducing Dynamic Drug Data examination by Hybrid Transductive Support Vector Machine with Fuzzy C Means (DDDA-HTSVM-FCM) procedure whose primary objective is to carry out drug tweet classification to acquire users' opinion. But the research lacks in terms of accuracy of opinion identification. Accuracy of classification might get diminished in prior work with the occurrence of more extraneous tweets aggregated in online. The classification recital might be lesser in the prevailing work with the arrival of huge volume of data with either labeled or unlabeled data. These crises are resolved in the anticipated research technique by introducing Relevancy and Similarity Aware Drug Comment Classification Framework (RSDCCF). In this research work, similarity-based filtering technique is provided to eliminate irrelevant reviews from a number of tweets accumulated online. It is performed by evaluating semantic similarity among verbal of sentences and score is allocated based on similarity level. Verbal with fewer score will be eradicated to raise learning performance and to diminish computation overhead which caused by processing irrelevant data. Then, Classification performance is enhanced with presence of mixed data by initiating ensemble classifier. Classifiers used in this work are Adaboost, SVM, Random Forest, and TSVM. The complete implementation of research technique is carried out in MATLAB simulation, and it is proven that anticipated technique offers optimal outcome associated with accurate prediction of user opinion in drug comments than prevailing work.

D. K. Renuka (✉)

Department of Computer Science (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India

B. R. Jeetha

Department of Computer Science, Dr. N.G.P College of Arts and Science, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_84

Keywords Drug posts · Twitter · Ensemble classification · Irrelevant post removal · Similarity calculation · Verbal score

Abbreviations

ADE	Adverse drug events
DDDA	Dynamic Drug Data examination
FAERS	FDA Adverse Event Reporting System
FCM	Fuzzy C Means
FDA	Food and Drug Administration
HTSVM	Hybrid Transductive Support Vector Machine
IRC	Internet Relay Chat
POS	Part-of-speech
RSDCCF	Relevancy and Similarity Aware Drug Comment Classification Framework
SVM	Support vector machine
TSVM	Transductive support vector machines
URL	Uniform resource locators

84.1 Introduction

Opinion mining is a procedure for analyzing and detecting subjective information, opinionated aspects, and sentiments in large volumes of text with computational methods [1]. Opinion mining is functional in various fields, especially to measure the product acceptance and services, along with the usual sentiments of people and brands [2]. This work spotlights on opinion mining in analyzing drugs, where patients have the rights to express their opinions and experiences about the medicines and the treatments. This domain application can attain a huge interest in current times. In specific, in pharma covigilance, manufacturers of drugs gain benefits from opinion mining as adverse effects to drug can be acquired more rapidly from posts or public repositories in social networks [3].

Adverse drug events (ADEs) are harms attained with the drug usage. Along with it, the purpose of being a foremost healthcare burden, ADEs raise the drug discovery costs and developments [4]. Indeed rigorous clinical and preclinical studies about the undiscovered ADEs cause drug withdrawals from market, yielding a significant loss in financial circumstances of pharmaceutical companies. However, it is very mandatory to identify and sense ADEs. The US Food and Drug Administration (FDA) is a regulatory agency charged with assuring safety for humans from drug usage. For postmarketing drug safety observation, FDA maintains a database known as FDA Adverse Event Reporting System (FAERS) [5]. Moreover, healthcare providers are

needed to reduce the reports of serious ADEs, and consumer submission is independent [6]. Therefore, numerous ADEs are not provided by FAERS. The social media websites explosion like Facebook, Google+, and Twitter offers consumers with an accessible channel to post their experiences with drug usage. As well, social media systems are often provided for receiving and soliciting feedback from consumers [7].

The goal of this investigation is to produce a computational pipeline to identify signals about ADEs by mining the opinions of users' in Twitter postings. Twitter was chosen as it offers an easy-to-use API streaming for getting real-time data. Meanwhile, extracting entire ADEs from tweets is probable, when adverse drug reactions (ADRs) were utilized as use case in this examination. ADEs are attained from the harm produced by using the drugs like overdose and ADRs, increase or decrease in dose or discontinuation of drug therapy.

Therefore, ADRs are side effects of using drugs that are caused by drug at normal use as normal dose.

In this investigation, drug opinion mining is executed to acquire knowledge about user viewpoint on specific drugs. Therefore, drug investigators can obtain decision about quality of drug efficiently. This foremost involvement of this work is to execute this framework, which can examine and accumulate drug tweets placed by diverse users to identify their opinions about fussy drug products. This investigational work guarantees specific calculation of analysis about drug-associated posts.

The complete research work structure is depicted as follows: In Sect. 84.1, introduction about the need of drug opinions and their role in drug production is discussed. In Sect. 84.2, varying associated research methodologies whose main goal is to perform the drug comment opinion are discussed. In Sect. 84.3, anticipated research methodologies with their working procedure and required samples are given. In Sect. 84.4, simulation evaluation part is discussed in terms of accuracy. Finally, in Sect. 84.5, overall conclusion of research work is provided by obtained simulation outcome.

84.2 Related Works

Nikfarjam and Gonzalez [8] proposed a technique for catching the fundamental syntactic and semantic examples from Web-based social networking posts. Machine learning-based methodologies have likewise been connected for ADR relation extraction and for classification of medications into classes, incorporating our pilot thinks in this field. But, there is still no examination that endeavors to perform deep linguistic investigations of Web-based social media text to create rich features. Gurulingappa et al. [9] utilized the openly accessible ADE corpus to perform binary classification utilizing different classifiers, acquiring a most extreme F-score of 0.77 for ADR class. Since this dataset is openly accessible, this gives us the chance of assessing our methodology on this information and against this benchmark framework.

Yeleswarapu et al. [10] consolidate client posted remarks from Internet-based media with information organized from databases and MED-LINE modified works and demonstrate that these sources can be used for ADR recognition. Freifeld et al. [11] present a methodology for utilizing client posts from Twitter ADR location utilizing a lexicon-based methodology. But, the methodology depends on the manual classification of ADR containing tweets. These exceptionally use online networking information and additionally spur the requirement for effective automatic classification approaches that are equipped for separating Web-based social networking content for ADR detection procedures.

Greene et al. [12] subjectively investigated the correspondences of Facebook group committed to diabetes. They found numerous advantages for patients taking an interest in these networks, for example, network assistance and way to learning, with proof of networks supporting unsafe practices; but, one fourth of posts are unequivocal ads, and some promote non-FDA (Food and Drug Administration)-affirmed items. Sixty-six percent of posts are portrayals of individual encounters in managing diabetes and one fourth of posts with delicate data probably not going to be uncovered in doctor–patient communications.

Denecke and Nejdil [13] examined different Medicine 2.0 substance and acquired patient-wrote a posting comprises medication associated ideas than some other post. However, Nejdil demonstrated that medication contains numerous illness-associated ideas and inferred that clients hunting for medications will discover outcomes in patient-composed posts. Lu et al. [14] contemplated substance of three sections, from online health community; they utilized one board for diabetes and two for disease. They understood medication-associated postings represented bigger fraction about board of diabetes than boards of cancer.

Weitzman et al. [15] examined quality and security of diabetes-associated OSNs and found quality/security of data was diverging over ten destinations during examination. Shrank et al. [16] likewise subjectively examined 15 diabetes-associated OSNs – of which comprises question or discussion gathering – and understood extensive range in individuals quality (from 3000 to 300,000), 33% OSNs gave doctors questions, and 66% are review posting. Zhang et al. [17] investigate about posts from Facebook community for diabetes and found more than 60% of posts give data, trailed by enthusiastic help (17%) and inspiring data (12%).

Goeriot et al. [18] manufactured and assessed sentiment lexicons utilizing drug surveys from social network. They fabricated lexicon dependent on prior lexicons from state of art, and an area dictionary dependent on medication reviews in social network. The author demonstrated opinion mining of health social media is conceivable, utilizing general mix and domain lexicons accomplishes finest outcomes. Bian et al. [19] constructed classifiers dependent on Twitter posts, former to anticipate if client (or somebody known) has utilized specific medication and latter to characterize if post portrays adverse medication event. Author acquires sensible exactness; however refer to noise in Twitter presents as one restriction on their methodology.

Chee et al. [20] anticipated whether medication pulled back through FDA posts in Yahoo! While anticipated classifier numerous false positives (as false positive), greater part of false positives with finest scores that pulled back from certain market

for timeframe. Yang et al. [21] utilized affiliation rule mining to recognize unfavorable medication in health social network. Utilizing FDA information, they affirmed relationships among medications and responses in posts. Leaman et al. [22] approved that client comments from social network are dug for unfavorable medication. They constructed lexicon dependent on comments of clients' posts to accomplish sensible precision utilizing lexical matching.

84.3 Relevancy and Similarity Aware Drug Comment Classification Framework

In this investigational work, drug opinion mining is executed to acquire knowledge about user viewpoint on specific drugs. Therefore, drug investigators can obtain decision about quality of drug efficiently. The foremost involvement of this work is to execute this framework, which can examine and accumulate drug tweets placed by diverse users to identify their opinions about fussy drug products. This investigational work guarantees specific calculation of analysis about drug-associated posts. The steps involved in this anticipated research framework for attaining drug opinion mining are given as:

- Primarily, drug tweet data are accumulated from social media website.
- The accumulated information is preprocessed to eliminate start and stop words and outcome with verbal words.
- Similarity score sourced filtering is to eradicate irrelevant data which are not associated with drug domain.
- At last, drug tweet classifications are performed to regain user opinion by initiating ensemble classifier – SVM, Adaboost, Random Forest, and TSVM.

The detailed elucidation of anticipated research method is provided in the following subsections.

Drug Tweet Gathering User-posted comments can be reclaimed from Twitter via MATLAB simulation vigorously. Twitter is a social networking website and online news where users and irrelevant messages, “tweets,” limited to 140 characters. Users registered can place tweets; however, those who have not registered can read them merely. Tweets are visibly seen by default, yet senders can restrict message to followers only. Users tweet through compatible external applications (like smartphones) and Twitter website or by short message service (SMS) accessible in definite countries. Users subscribe to other users – this is termed “following,” and “followers” or “tweets” are known as subscribers, peeps, and Twitter. Individual tweets are endorsed by users of own feed, process termed as “re-tweets.” Users may “like” (usually “favorite”) individual tweets. It facilitates users to update profile through mobile either by apps released for certain tablets and smartphones or text messaging. Twitter in contrast to Web-based Internet Relay Chat (IRC) client. Time

essay technologies founder Steven Johnson explained essential Twitter mechanisms as “remarkably simple”: Twitter links are user post and downloaded as “https://twitter.com/master_usmle”.

Preprocessing Drug Tweet Dataset Twitter is a microtext social networking website. A tweet typically comprises hashtags, uniform resource locators (URLs), misspelled words, emoticons, incorrect grammar expressions, and Internet slang. Python’s regular built-in expressions were utilized to progress text in tweet. The entire re-tweets and non-English tweets (tweets forwarded or reposted by other users) were eliminated. As well, tweets, which have “e-book” in hashtags, were eliminated. As certain advertisement or spam accounts are usually connected to medical documents, they generally incorporate “e-book” in hashtag. Indeed, the following measures were considered to de-noise tweets:

- Duplicate tweets were eliminated.
- Hashtags and usernames were restored with space.
- Added punctuation characters (except “!” and “?”) and newline (\n) were eliminated.

Subsequently, stemming, tokenization, and part-of-speech (POS) mining were carried out. Tokenization refers contravention sentence into words. POS is hauling out of lexical items that have same grammatical properties. Stemming means the procedure for decreasing inflected words to root forms. For instance, “ate,” “eating,” “eats,” and “eaten” are stemmed into “eat.” Text Blob, Python library for processing textual information, was utilized to execute these tasks.

Similarity Score-Based Filtering Similarity score-based filtering is functional to eliminate irrelevant data which is not associated with drug domain. It is performed by evaluating semantic similarity among verbal of sentences, and score is allocated based on the similarity level. The verbal with less score can be eliminated to increased learning performance and diminish computation overhead which occurs during processing irrelevant data.

The composition of the entire words in a sentence carries concept and meaning of sentence. Therefore, sentence which has independent and complete meaning may carry full semantic load. The semantic load of entire sentence is shared between its parts. Distinctive of functional words, content words have primary role to hold the sentence meaning. But, how sharing of each word for semantic load of entire sentence is carried out? It is evident that it cannot evaluate an equal value for each word’s semantic load share. Some sentence part carries greater portion of meaning for sentence to other parts. For instance, it is not suitable that for every sentence like “I take tablets regularly,” each word has similar value 0.25. Evidently, determiner here holds lowest sentence portion of semantic load. If determiner is eliminated from the sentence, the foremost idea will be clear. If the word “take” is omitted from the sentence, the outcome of the sentence is “I tablets recurrently” would be meaningless; hence, meaning cannot be understand from it. If word like “tablets” is eliminated from the sentence, the outcome sentence will be “I take regularly” is unfinished and unclear, but reading action is inferable from it. It has to be noted

that phrasal and compound verbs should be measured atomically, i.e., prepositions have not to be separated from significant part of verb. For instance, preposition “off” cannot be divided from verb “take off.”

Owing to this verb of certain sentence holds significant semantic portion that will be approved in future; the anticipated technique is grounded on partitioning sentence into three parts: *verb, before, and after verb* order. Every partition holds a share of *semantic load* of two sentences, as sum of similarity level among three subsequent parts in two sentences. To evaluate similarity of two sentences, every three parts in two sentences are coordinated and contrasted likewise; then similarity among every two corresponding parts is computed based on diverse matching levels. After that three measured similarity values are incorporated to evaluate whole similarity of the investigation.

If S1 and S2 are two sentences in which similarity value has to be measured, initially, every sentence is divided into three steps, as trails:

S1: {m words before Verbs1} {Verbs1} {n words after Verbs1} $m, n \geq 0$

S2: {p words before Verbs2} {Verbs2} {q words after Verbs2} $p, q \geq 0$

Stem (Verbs1) = Stem (Verbs2)

The foremost significant part of matching procedure is to categorize verbs. Before dividing a sentence, identify POS tags of each word in sentence. POS-tagger component recognizes appropriate POS of verbs that may have Noun POS with certain dependencies such as determiners and prepositions occur near verb-form word. If input sentence is:

(a) I got headache after taking fever tablet

Then tagged sentence will be as below; the word *copy* is recognized as noun rather than verb:

I < N > got<V > headache<PN> after<AD> taking<V > fever<Prep> tablet<Prep>

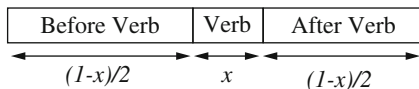
and five divisions of this sentence will be:

I | got headache | after | taking | fever tablet

Similarity measure is a number that lies between 0 and 1. Value 1 points exact similarity, while 0 refers to complete difference of sentences. When similarity measure is nearer to 1, two sentences will be more similar. Similarity measure would be shared among diverse parts of sentence that has to answer. There is not any precise or estimated measure of semantic load shared between sentence parts. As an effortless heuristic, measure an equal semantic load share for before-verb and after-verb parts. Therefore, determine semantic load share for verb part. As depicted in Fig. 84.1, if x is a share of semantic load for verb part, then share semantic load for either before- or after-verb parts will be $(1-x)/2$. Figure 84.1 illustrates partitioning sentence and similarity score related to every part.

Based on allocated score, filtering will be performed where verbs with lesser score will be eradicated for user opinion mining procedure.

Fig. 84.1 Sentence partitioning and semantic load sharing among partitions



Drug Tweet Classification In machine learning and statistics, ensemble techniques make use of numerous learning procedures to acquire superior predictive recital that can be acquired from constituent learning procedures. Contrasting to statistical ensemble in mechanics that is typically infinite, machine learning ensemble specifies concrete finite set of alternative replicas; however, it usually permits much flexible structure to subsist alternatives. In this investigation, this ensemble classification algorithm is adapted to acquire superior prediction precision based on accurate identification of user opinion. This is attained by initiating weighted-based ensemble process which tries to merge the outcomes of diverse classifiers with weight based on which accurate analysis would be performed.

A novel weighted voting classification ensemble technique is presented and is termed as WAVE which utilizes two weight vectors: instances weight vector and classifiers weight vector. Instance weight vector allocates greater observations weight that is complex to categorize.

Classifiers weight vector places higher weights over classifiers that execute finest on hard-to-classify samples. Weight vector is intended to conjunction computation with other by iterative process. Instances of huge weights perform more significant role in influencing classifiers weight and vice versa. It is depicted that iterated weight vectors unite to optimal weights that are openly computed from classifiers performance matrix in ensemble. Ultimate ensemble prediction is attained by voting by optimal classifiers weight vector. Classifiers considered ensembling are “Random Forest, SVM, Adaboost, and TSVM.”

Support Vector Machine Classification SVM is supervised learning replicas with associated learning algorithms that examine information and identify patterns, utilized for regression analysis and classification. Provide training sample sets, every mark for fits into two groups; an SVM training procedure produces model that allocates novel samples into one group or other.

Formula

To diminish error minimization, use the formula below:

$$Dt + 1(i) = Dt(i) \cdot e^{ytMt(xi)} \tag{84.1}$$

$$\Phi(w) = \frac{1}{2} \|w\|^2 \tag{84.2}$$

Approximating function

$$F(x) = \sum_{i=1}^{nsv} (x_i, y_i) \quad (84.3)$$

Algorithm

Provided dataset $X = (x_1, y_1), \dots, (x_n, y_n)$, $C // x$ and y – labeled sequence and C-class

Initialize vector $v = 0$, $b = 0$; class // v-vector and b-bias

Train preliminary SVM

For every $x_i \in X$ do // x_i is vector comprising features determines sample i

Categorize x_i by $f(x_i)$

If $y_i \neq f(x_i) < 1 //$ prediction class label

Acquire w' , b' for known information // w' , b' for novel data

Add x_i to identified data

If prediction is incorrect then retrain

Repeat

End

Error esteems are reduced by Eq. (84.1)

Categorize outcomes by Eq. (84.2)

Adaboost Classification AdaBoost (Adaptive Boosting) was initially developed by Schapire and Freund (1996), which is most admired ensemble procedure that enhances boosting procedure through an iterative procedure. Foremost concept of this process is to provide added potential to complex patterns to classify. Focal point is enumerated by weight is allocated to each pattern in training set. At first, similar weight is allocated to all patterns. In every iteration weights of entire misclassified samples are maximized, while properly classified samples weights are reduced. However, weak learner is enforced to focus on difficult training set samples by performing added iterations and producing added classifiers. As well, weight is assigned to every individual classifier. This weight measures complete classifier precision and total weight of function suitably classified patterns. Consequently, higher weights are offered to more accurate classifiers. These weights are cast off for classification of patterns. This iterative procedure gives classifiers series that harmonize one another. It has been demonstrated that AdaBoost approximates margin classifier such as SVM.

Algorithm

Prerequisites: I (a weak inducer), T (the number of iterations), S (training set).

Guarantee: $M_t, \alpha_t; t = 1 \dots T$

1. $t \leftarrow 1$
2. $D_t(i) \leftarrow 1/m; i=1,\dots,m$
3. Repeat
4. Generate classifier M_t by I and distribution D_t
5. $E_t \leftarrow \sum_{i: M_t(x_i) \neq y_i} D_t(i)$
6. If $E_t > 0.5$ then
7. $T \leftarrow t-1$
8. Exit loop
9. End if
10. $\alpha_t \leftarrow \frac{1}{2} \ln \left(\frac{1-E_t}{E_t} \right)$
11. $D_{t+1}(i) = D_t(i) \cdot e^{-\alpha_t y_i M_t(x_i)}$
12. Normalize D_{t+1} as appropriate distribution
13. $T++$
14. Until $t > T$

Random Forest Classification Random Forest ensemble (random subspace) makes use of un-pruned decision trees, huge amount of individual. Individual trees are generated with algorithm. The input factor N specifies sum of input variables that is utilized to describe decision at tree nodes. This number has to be lesser than amount of attributes in training set. Bagging can be considered as Random Forests cases attained while N is attributes amount in unique training set.

Algorithm

Prerequisites: T (the number of iterations), IDT (decision tree inducer), μ (the subsample size), S (the training set), and N (sum of attributes cast off in each node).

Guarantee: $M_t; t = 1,\dots,T$

1. $t \leftarrow 1$
2. repeat
3. $S_t \leftarrow \mu$ instances sample from S with replacement
4. Design classifier M_t by $IDT(N)$ on S_t
5. $t++$
6. Until $t > T$

Transductive Support Vector Machine Transductive support vector machines (TSVM) have been widely cast off as handling partly labeled data in semi-supervised learning.

In sense, supervised learning, transduction or transductive interface, statistical inference, is causing from certain (training) to explicit (test) cases. Similarly, inductions are reasoning from noticed training cases to common rules that are functional to test cases. Distinction is foremost interesting cases where transductive model prediction is not attainable by every inductive model. This is sourced by transductive inference on diverse test sets generating mutually conflicting predictions. Here

classification procedures, drugs mentions in user posts are recognized. As well, recently identified drugs also identify and it will be modified in drug dictionary for reference.

$$\text{Reduce over } (y_1^*, \dots, y_n^*, \vec{w}, b) : \frac{1}{2} \|\vec{w}\|^2$$

$$\text{Subject to } \bigvee_{i=1}^n : y_i [\vec{w} \cdot \vec{x}_i + b] \geq 1$$

$$\bigvee_{j=1}^k : y_j^* [\vec{w} \cdot \vec{x}_j + b] \geq 1$$

The foremost objective of TSVM is to attain test dataset labeling “Y” and $\langle w, b \rangle$ hyper plane; therefore this hyperplane divides both testing and training data with margin.

Weighted Voting Classification Ensemble While ensemble classifiers are generated, anticipated technique allocates distinctive voting 43 weights to every classifier belongs ensemble. Specially, by an iterative procedure, 44 weight vectors for classifiers and subsequent weight vector of samples is 45 attained in learning model formation stage. After final iteration, hard-to-classify samples attain greater 47 weights; consequently, enhanced carry out classifiers on hard-to-classify in 48 samples are allocated higher weights. Last computation of ensemble is attained by voting classifiers weight vectors.

Assume if there exist two classifiers. In mainstream voting, entire classifiers possess equivalent weights. Therefore, if two classifiers make diverse predictions for the same sample, ultimate decision turns out to be subjective owing to its tied votes. Currently, suppose initial one to categorize more hard-to-classify samples appropriately, while subsequent classifier makes simpler samples appropriately. If two classifiers generate diverse predictions on unnoticed samples, it is sensible to offer weight to preliminary classifier which splits more complex samples properly.

Iterative Weight Adjust Algorithm

1. Place weight vector instance initially

$$Q_0 = \frac{(J_{nk} - X)(J_{kk} - I_k)1_k}{1'_n(J_{nk} - X)(J_{kk} - I_k)1_k}$$

Q0 acquires greater weights for rows instance of “X” with lesser 1’s (i.e., hard-to-classify samples). Denominator is factor for normalizing unit norm.

2. For $m = 1, 2, \dots$ Replicate (a) and (b).
 - (a) Compute weight vector of classifier

$$P_m = \frac{X'Q_{m-1}}{1'_k X'Q_{m-1}}$$

P_m allocate greater weights on further precise classifiers (“X” columns more 1’s) after weight Q_{m-1} integration instance. Denominator is factor for normalizing unit norm.

- (b) Revise weight vector instance

$$Q_m = \frac{(J_{nk} - X)(J_{kk} - I_k)P_m}{1'_n (J_{nk} - X)(J_{kk} - I_k)P_m}$$

Q_m allocates greater weights on hard-to-classify samples after integrating classifier weight vector P_m . Denominator is factor for normalizing unit norm.

3. Discontinue step 2 while weight vectors P_m and Q_m turn to constant. Specify P^* and Q^* be ultimate weight vectors.

84.4 Experimental Results

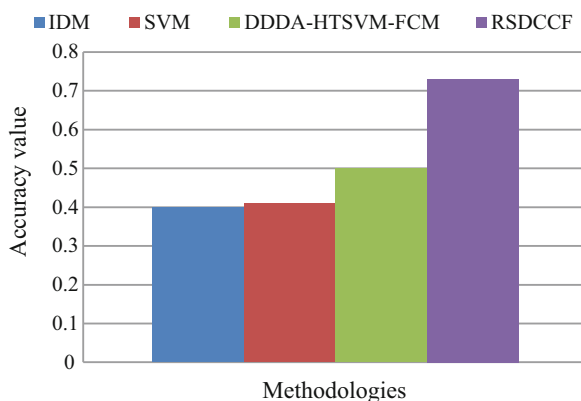
In this segment, numerical computation of anticipated research methodology is carried out based on diverse performance measures to examine performance improvement of anticipated and prevailing research methodologies. MATLAB simulation environment is utilized to execute anticipated research method. Performance measures assumed in this investigated are given as trails: “sensitivity, accuracy, specificity, and precision.” Proposed research technique is executed in MATLAB simulation environment which is examined for its recital over performance metrics. Comparison is made among anticipated method Relevancy and Similarity Aware Drug Comment Classification Framework (RSDCCF), prior method DDDA-HTSVM-FCM, and prevailing methodologies SVM and IDM.

Accuracy

It is depicted as degree of appropriate identification of drug names that occur in tweets, i.e., less false-positive rate. Accuracy of anticipated system should be greater than other prevailing techniques like IDM, DDDA-HTSVM-FCM, and SVM. Accuracy value is measured in terms of drug prediction system’s false-positive, true-positive, true-negative, and false-negative values. Accuracy is measured as follows (Table 84.1):

Table 84.1 Accuracy comparison values

Method	IDM	SVM	DDDA-HTSVM-FCM	RSDCCF
Accuracy	0.4	0.41	0.5	0.73

Fig. 84.2 Accuracy comparison**Table 84.2** Sensitivity comparison values

Methods	IDM	SVM	DDDA-HTSVM-FCM	RSDCCF
Sensitivity	0.659	0.7	1	1

$$\text{Accuracy} = \frac{T_p}{(T_p + F_p + F_n)}$$

Figure 84.2 depicts accuracy metric comparison with graphical representation. This graph illustrates that anticipated research techniques is superior than prevailing research techniques. RSDCCF acquires 46% enhanced accuracy than DDDA-HGSO-FSO-FSVM, 78% superior than IDM, and 82.5% superior than SVM.

Sensitivity

It is termed as true positive rate which is distinct as appropriate degree classifying drug data items as positive. Sample of medical test utilized to recognize disease, sensitivity of test is ratio of people who test positive for disease between those who have disease. Mathematically, this can be illustrated as (Table 84.2):

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}}$$

Figure 84.3 depicts sensitivity metric comparison with graphical representation. This graph illustrates that anticipated investigation is superior than prevailing research techniques. RSDCCF shows similar sensitivity like DDDA-HTSVM-FCM, 42.8% superior to IDM, 51.7% superior to SVM.

Fig. 84.3 Sensitivity comparison

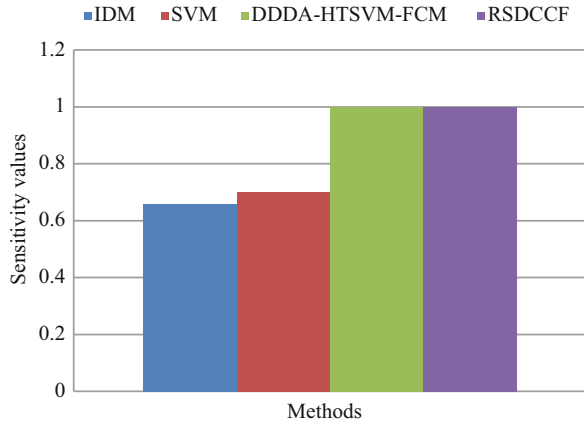


Table 84.3 Specificity values

Methods	IDM	SVM	DDDA-HTSVM-FCM	RSDCCF
Specificity	0.04	0.045	0.0522	0.071

Specificity

It is termed as true-negative rate which is depicted as appropriate degree which classifies non-drug data items. Assume an instance of medical test for recognizing disease. Test specificity is proportion of healthy patients not to possess disease, who test negative for it. Mathematically, this can also be formulated as (Table 84.3):

$$\text{Specificity} = \frac{\text{number of true negatives}}{\text{Number of true negatives} + \text{Number of false positives}}$$

Figure 84.4 depicts specificity metric in comparison with graphical representation. This graph illustrates that anticipated research technique is superior than prevailing research techniques. RSDCCF attains 36% better specificity than DDDA-HTSVM-FCM, 57.7% better than IDM, and 77.5% superior than SVM.

Precision

It is proportion of instances retrieved that are relevant (Table 84.4)

$$\text{Precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Figure 84.5 depicts precision metric evaluation of graphical representation. This graphs shows that anticipated research technique is superior than prevailing research techniques. RSDCCF attains 37.88% better specificity than DDDA-HTSVM-FCM, 106% better than IDM, and 116% better than SVM.

Fig. 84.4 Specificity comparison

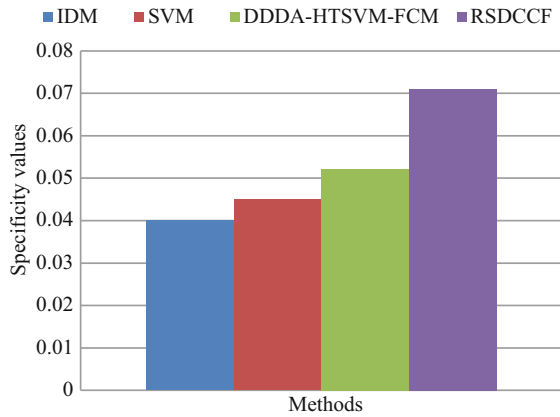
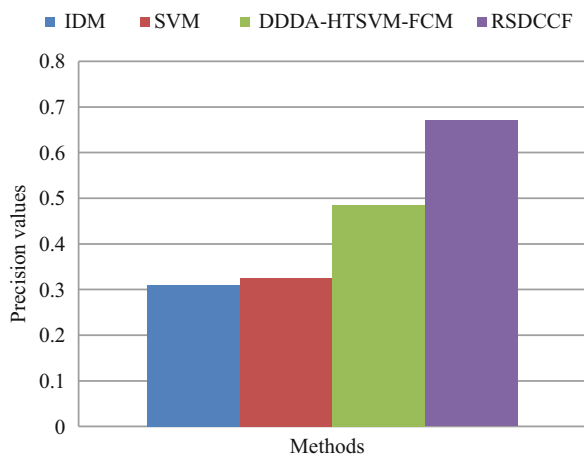


Table 84.4 Precision comparison values

	IDM	SVM	DDDA-HTSVM-FCM	RSDCCF
Precision	0.31	0.325	0.4859	0.67

Fig. 84.5 Precision comparison



84.5 Conclusion

In this investigation, similarity-based filtering method is anticipated to eliminate irrelevant reviews from a number of tweets aggregated online. It is performed by measuring semantic similarity among verbal of sentences and score is allocated based on similarity level. Verbal with fewer score will be eliminated to raise learning performance and decrease computation overhead which might happen by processing

irrelevant data. Classification performance is enhanced with the occurrence of mixed data by introducing ensemble classifier. These classifiers are used in this work Random Forest, SVM, Adaboost, and TSVM. The complete implementation of this research technique is performed in MATLAB simulation environment from it is depicted that anticipated research technique helps to offer optimal outcome in association with accurate calculation of user opinion on drug comments than the prevailing work.

References

1. Serrano-Guerrero J, Olivás JA, Romero FP, Herrera-Viedma E (2015) Sentiment analysis: a review and comparative analysis of web services. *Inf Sci* 311:18–38
2. Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowl-Based Syst* 89:14–46
3. Khan FH, Bashir S, Qamar U (2014) TOM: twitter opinion mining framework using hybrid classification scheme. *Decis Support Syst* 57:245–257
4. Mann E, Zepeda O, Soones T, Federman A, Leff B, Siu A, Boockvar K (2018) Adverse drug events and medication problems in “Hospital at Home” patients. *Home Health Care Serv Q*:1–10
5. Food and Drug Administration, HHS (2016) Deeming tobacco products to be subject to the Federal Food, drug, and cosmetic act, as amended by the family smoking prevention and tobacco control act; restrictions on the sale and distribution of tobacco products and required warning statements for tobacco products. Final rule. *Fed Regist* 81(90):28973–29106
6. Peddie D, Small SS, Badke K, Bailey C, Balka E, Hohl CM (2018) Adverse drug event reporting from clinical care: mixed-methods analysis for a minimum required dataset. *JMIR Med Inform* 6(2):e10248
7. Heimbach I, Schiller B, Strufe T, Hinz O (2015) Content virality on online social networks: empirical evidence from Twitter, Facebook, and Google+ on German news websites. In: *Proceedings of the 26th ACM conference on hypertext & social media*. ACM, New York, pp 39–47
8. Nikfarjam A, Gonzalez G (2011) Pattern mining for extraction of mentions of adverse drug reactions from user comments. In: *Proceedings of the American Medical Informatics Association (AMIA) annual symposium*, pp 1019–1026
9. Gurulingappa H, Fluck J, Hofmann-Apitius M, Toldo L (2011) Identification of adverse drug event assertive sentences in medical case reports. In: *First international workshop on knowledge discovery and health care management (KD-HCM), European conference on machine learning and principles and practice of knowledge discovery in databases (ECML PKDD)*, pp 16–27
10. Yeleswarapu S, Rao A, Joseph T, Saipradeep VG, Srinivasan R (2014) A pipeline to extract drug-adverse event pairs from multiple data sources. *BMC Med Inform Decis Mak* 14(13):1–16
11. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, Dasgupta N (2014) Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf* 37(5):343–350
12. Greene JA, Choudhry NK, Kilabuk E, Shrank WH (2011) Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *J Gen Intern Med* 26(3):287–292
13. Denecke K, Nejdil W (2009) How valuable is medical social media data? Content analysis of the medical web. *Inf Sci* 179(12):1870–1880
14. Lu Y, Zhang P, Liu J, Li J, Deng S (2013) Health-related hot topic detection in online communities using text clustering. *PLoS One* 8(2):e56221

15. Weitzman ER, Cole E, Kaci L, Mandl KD (2011) Social but safe? Quality and safety of diabetes-related online social networks. *J Am Med Inform Assoc* 18(3):292–297
16. Shrank WH, Choudhry NK, Swanton K, Jain S, Greene JA, Harlam B, Patel KP (2011) Variations in structure and content of online social networks for patients with diabetes. *Arch Intern Med* 171(17):1589–1591
17. Zhang Y, He D, Sang Y (2013) Facebook as a platform for health information and communication: a case study of a diabetes group. *J Med Syst* 37(3):1–12
18. Goeuriot L, Na JC, Min Kyaing WY, Khoo C, Chang YK, Theng YL, Kim JJ (2012) Sentiment lexicons for health-related opinion mining. In: *Proceedings of the 2nd ACM SIGHIT international symposium on health informatics*. ACM, New York, pp 219–226
19. Bian J, Topaloglu U, Yu F (2012) Towards large-scale twitter mining for drug-related adverse events. In: *Proceedings of the international workshop on Smart health and wellbeing*. ACM, New York, pp 25–32
20. Chee BW, Berlin R, Schatz B (2011) Predicting adverse drug events from personal health messages. In *AMIA annual symposium on American medical informatics association*, 217–26
21. Yang CC, Jiang L, Yang H, Tang X (2012) Detecting signals of adverse drug reactions from health consumer contributed content in social media. In *Proceedings of ACM SIGKDD workshop on health informatics*, pp 1–8
22. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G (2010) Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. In *Proceedings of the workshop on biomedical natural language processing*, pp 117–125

Chapter 85

Enhanced Particle Swarm Optimization with Genetic Algorithm and Modified Artificial Neural Network for Efficient Feature Selection in Big Data Stream Mining



S. Meera and B. Rosiline Jeetha

Abstract High dimensionality would be one of the major challenges faced by people working in research with big data as a high dimensionality that happens, while a dataset comprises of a big number of features. For resolving this issue, often researchers make use of a feature selection step for identification and removal of irrelevant features and repetitive features. Acceleration Artificial Bee Colony-Artificial Neural Network (AABC-ANN) has been introduced in the preceding research for handling the feature selection process over the big data. Computational complexity and inaccuracy of dataset remain as a problem for these methods. Enhanced Particle Swarm Optimization with Genetic Algorithm – Modified Artificial Neural Network (EPSOGA–MANN) is described in the proposed methodology for avoiding the above-mentioned issues. Modules including preprocessing, feature selection, and classification have been included in this research process. Fuzzy C Means (FCM) denotes the clustering algorithm which is used to handle the noise information efficiently in preprocessing. Feature selection process is carried out by means of EPSOGA algorithm optimally in this research. More important and relevant features are selected by EPSOGA optimization algorithm and as a result more accurate classification results are achieved in this work for huge volume of dataset. Input, hidden, and output layers are the three layers of MANN. It is introduced for improving the time complexity by means of neurons. The performance evaluation of the research method is conducted in the Matlab simulation environment.

S. Meera (✉)

Department of Computer Science (PG), PSGR Krishnammal College for Women, Coimbatore, Tamil Nadu, India

B. R. Jeetha

Department of Computer Science, Dr. N.G.P College of Arts and Science, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_85

909

Keywords Feature selection · Classification · Enhanced Particle Swarm Optimization with Genetic Algorithm (EPSOGA) · Modified Artificial Neural Network (MANN) · Big data

Abbreviations

AABC-ANN	Acceleration Artificial Bee Colony-Artificial Neural Network
ACO	Ant Colony Optimization
EPSOGA-MANN	Enhanced Particle Swarm Optimization with Genetic Algorithm – Modified Artificial Neural Network
FCM	Fuzzy C Means
GA	Genetic Algorithm
SA	Simulated Annealing
SHARP	Stream-based Holistic Analytics and Reasoning in Parallel

85.1 Introduction

Data generation from various sources takes place in the digital world and the rapid move from digital technologies has resulted in the development of big data. These are seen in structured, semistructured, and unstructured format in the size of petabytes and more [1]. In formal it is specified from 3 Vs to 4 Vs. 3 Vs and defines volume, velocity, and variety [2]. Huge volumes of transactional, sensory, and Web data on everyday basis are consistently created in the form of streams, which need to be evaluated online on their arrival. Data streaming is considered to be one among the major resources of big data [3]. Lot of attention has been paid over the last decade for predictive modeling of data streams and big data, various research approaches are designed for the predictive modeling of data streams and big data for well-built controlled problem configuration, on overlooking significant issues possessed by practical applications.

Important data processing step is Feature Selection (FS) which is used extensively in the machine learning and data analytics problems [4]. Many algorithms are available for clustering and feature selection process. Hence this big data must involve the appropriate algorithm. For finding the optimal feature subset, Random search [5] along with different approaches has been employed in various feature selection methods. The search process proceeds in two different variations as it starts with a randomly selected feature subset [6]. Based on the current subset, new subsets are being generated by instance-based methods, and they use heuristics for generating and updating the subsets. Exemplar of the algorithms, which are instance-based, includes tab search [8], Simulated Annealing (SA) [7], and Genetic Algorithm (GA) [9]. Other variations, i.e., methods based on the model, basically rely on the distribution of sample and the update parameters of probability distribution. Ant

Colony Optimization (ACO) [10, 11] belongs to few examples of this category, which is stochastic gradient search. Computational effort becomes high for the trade-off in some situation.

85.2 Literature Survey

A distributed adaptive model procedure used for the mining of big data streams is designed by Vu et al. [12]. A distributed streaming procedure is initially designed for learning the decision rules for tasks involving regression. Peralta et al. [13] designed an evolutionary feature selection for big data classification. It makes use of the MapReduce paradigm to acquire the subsets of features out of big datasets. Xin et al. [14] designed a flexible extreme learning machine that depends on MapReduce framework, known as Elastic ELM (E²LM), to surpass the drawbacks of ELM* whose learning capability is not quite strong when used for the updated big-scale training dataset. Fong et al. [15] implemented a new Scalable Data Stream Mining approach in big data. Here, in this system, a new data stream mining technique, known as Stream-based Holistic Analytics and Reasoning in Parallel (SHARP), is introduced.

Fong et al. [16] introduced a new feature selection procedure for data stream mining of big data. Feature selection has been generally utilized to reduce the processing load in the induction of a data mining model. In order to reduce high dimensionality and the streaming behavior of data in big data, a new compact feature selection is introduced. The feature selection is modeled especially for mining the streaming data on the run, by making use of the Accelerated Particle Swarm Optimization (APSO) kinds of swarm search, which accomplishes the improved analytical correctness within the acceptable processing time. This methodology also fits in a good manner with current applications anywhere their data reach in streams.

85.3 Proposed Methodology

85.3.1 Dataset Collection

HIGGS dataset has been acquired from UCI archive¹ for simulating the negative impacts of the Big Data in accordance with high dimensions (several features) and massive volume (several instances) for the purpose of experimentation. It is a physics dataset of particle detectors in accelerator. The data is being generated by means of Monte Carlo simulations. Almost 53% of positive samples are used for balancing the dataset. The dataset consists of 11,000,000 samples. The HIGGS dataset features are described in Table 85.1.

Table 85.1 Details of HIGGS dataset

Feature name	Feature details
Lepton pT, lepton eta, lepton phi, missing energy magnitude, missing energy phi, jet 1 pt., jet 1 eta, jet 1 phi, jet 1 b-tag, jet 2 pt., jet 2 eta, jet 2 phi, jet 2 b-tag, jet 3 pt., jet 3 eta, jet 3 phi, jet 3 b-tag, jet 4 pt., jet 4 eta, jet 4 phi, jet 4 b-tag	21 low-level features. These form kinematic attributes
m_jj, m_ijk, m_lv, m_jlv, m_bb, m_wbb, m_wbbb	7 high-level features employed to distinguish between the two classes

85.3.2 Fuzzy C Means (FCM) Clustering Algorithm

In preprocessing, for efficient handling of noisy data, Fuzzy C Means (FCM) clustering algorithm is used. Aggregation of the n data points $x_i, i = 1 \dots n$ is partitioned by FCM into c fuzzy groups which gets a cluster center in every group so that the dissimilarity measure objective function is reduced. The membership matrix U is let to possess elements having values between 0 and 1. But, the sum of the degrees of belongingness is stipulated on imposing normalization for a dataset that always has to be equivalent to unity, as expressed in Eq. (85.1):

$$\sum_{i=1}^c u_{ij} = 1, \forall_j = 1 \dots n \tag{85.1}$$

Summation of squared distances between the cluster centers and the points that belong to these clusters multiplied by the membership values (grades) of every cluster generated by the FCM is represented by the objective function (Euclidean). The objective function for FCM is:

$$J(U, c_1, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_j u_{ij}^m d_{ij}^2 \tag{85.2}$$

Where,

u_{ij} – Membership of x_i in the cluster j , values between 0 and 1

c_i – Cluster center of fuzzy group i

$d_{ij} = \|c_i - x_j\|$ – Euclidean distance between i th cluster and j th data point

$m \in (1, \infty)$ – weighting exponent (any real number more than 1)

The required conditions for Eq. (85.2) to attain its minimum are:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \tag{85.3}$$

and

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{85.4}$$

The FCM algorithm is considered to be a simple iteration process through the preceding two required conditions. The algorithm always decides the cluster centers c_i and the membership matrix U by utilizing the steps that follow [17]:

Step [1]: Initialization of the membership matrix U with the arbitrary values between 0 and 1 so that the conditions in Eq. (85.1) will be met.

Step [2]: Compute c fuzzy cluster centers $c_i, i = 1 \dots c$, by means of Eq. (85.3).

Step [3]: Computation of the objective function as per Eq. (85.2). Stop the process if either of it is less than a specific tolerance value or its enhancement over the earlier iteration is less than the particular threshold value, ϵ .

Step [4]: Calculate a new U utilizing Eq. (85.4). Go to Step 2.

At final stage, the noisy, missing, and error data have been grouped and removed from dataset.

85.3.3 Enhanced Particle Swarm Optimization with Genetic Algorithm Based Attribute Selection

General GA Algorithm

Genetic Algorithm (GA) is considered to be an important class of an evolutionary algorithm. Approximate solutions have been usually provided by GA to the various problems. The cost function is usually considered to be the objective function of the given problem. Some processes associated with GA are:

- (a) Selection – this process in general is used for choosing the chromosome for reproduction based on the fitness criterion.
- (b) Reproduction – this step is involved for the formation of next generation from the current generation.
- (c) Crossover – this process is used for exchanging genetic material between chromosomes.
- (d) Mutation – process which leads to the change in chromosomes for an individual.
- (e) Stopping criteria – this process is the last step in GA. When the desired solution is reached, the iteration stops or its maximum number of cycles is achieved.

General PSO Algorithm

In order for identification of the finest solution, a set of particles searches through the solution space in PSO. Every solution found in particle swarm optimization is represented as a discrete bird in the swarm, such that a particle is in the search space. $X_i = (x_1, x_2, \dots, x_d)$ is the location of the i th particle and in D -dimensional search space, velocity V_i will be denoted as (v_1, v_2, \dots, v_d) . The various steps used in the PSO algorithm are given below:

- (a) Initialization of the particles with some arbitrary velocities and the positions in search space.

- (b) Corresponding value of fitness function has been calculated for the swarm particles.
- (c) Evaluation of fitness value is equated with the current value of the particle's pbest. In comparison with pbest, when the current value is better, fix the value as new pbest value and fix the pbest location as the present location in the n-dimensional space.
- (d) Equate the fitness value as a next with the previous overall best value. Reset the gbest in case the present value is better compared to gbest to the array value of the current particle and index.
- (e) Finally, these values are assigned to the corresponding velocity and position of the swarm particle.
- (f) End the process while the criteria are satisfied.

Hybrid Approach

The GA and PSO is a stochastic search process which is population based. For feature selection, the system proposes Enhanced Particle Swarm Optimization, a new technique with Genetic Algorithm-Artificial Neural Network (EPSOGA). Population of size $2N$ is randomly initialized by EPSOGA. The crossover of P_{id} and P_{gd} occurs in enhanced PSO for each particle. Two child particles could be gained. Assume the particle with smaller fitness value to be considered as the last child particle after comparison. Thus, improved PSO could make use of better performing P_{id} and P_{gd} to lead all of the particles for searching the optimum solution. Thus, the Vid has been updated in accordance with Eq. (85.5):

$$V_{id}^{new} = W \times V_{id}^{old} + c_1 \times rand_1 \times (P_{id} - X_{id}) + c_2 \times rand_2 \times (P_{gd} - X_{id}) \quad (85.5)$$

and mutated. Then, X_{id} can be obtained using Eq. (85.6) as shown below:

$$X_{id}^{new} = X_{id}^{old} + V_{id}^{new} \quad (85.6)$$

EPSOGA Computational Procedures

The computational processes for the newly introduced EPSOGA are given as below:

- Step 1:** Decide on parameters that include the population size $2N$ (number of particles), inertia weight, W , maximum velocity, V_{max} , and two learning factors, c_1 and c_2 .
- Step 2:** Initialize every particle in random with the initial position, X_{id} and velocity, V_{id} .
- Step 3:** Compute fitness value (accuracy) for every particle so as to refine the suitability.
- Step 4:** Use real-coded GA operators (reproduction, crossover, and mutation) to the $2N$ population and generate another $2N$ population.
- (4a)Reproduction:** From the population, choose the $2N$'s best individuals based on fitness.
- (4b)Crossover:** Use two-parent crossover for updating the $2N$ individuals.

(4c)Mutation: Use mutation with 5% mutation probability to the 2 N chromosomes as per the equation given below:

$$x' = x + \text{rand} \times N(0, 1) \quad (85.7)$$

Step 5: Assess the fitness of each one of the 2 N individuals. Sort them based on the fitness values.

Step 6: Update the global best (P_{gd}) and local best positions (P_{id}).

Step 7: Crossover: On the basis of NPSO [39], every P_{id} and P_{gd} crossover is applied and two child particles are obtained, and compared and let the particle having the lesser fitness value be termed the final child of the predecessors.

Step 8 (8a) Update the velocity (Vid) applying Eq. (85.8).

$$V_{old}^{new} = W \times V_{id}^{old} + c_1 \times \text{rand}_1 \times (P_{id} - X_{id}^{old}) + c_2 \times \text{rand}_2 \times (P_{gd} - X_{id}^{old}) \quad (85.8)$$

Where

c_1 and c_2 – two positive constant

rand_1 and rand_2 – two random functions in the range [0, 1]

W refers to the inertia weight.

(8b) Mutation: Update the Vid using Eq. (85.9)

$$V_{old}^{new'} = V_{old}^{new} + \text{rand} \times N(0, 1) \quad (85.9)$$

Step 9 (9a) Update the position vector (Xid) of the topmost N individuals utilizing Eq. (85.10).

$$X_{old}^{new} = X_{id}^{old} + V_{id}^{new} \quad (85.10)$$

(9b) Mutation: Update the Xid of the topmost N individuals using Eq. (85.11).

$$X_{id}^{new'} = X_{id}^{old} + \text{rand} \times N(0, 1) \quad (85.11)$$

Step 10 Use K-means operators to the 2 N population and generate another 2 N population applying Eq. (85.12).

$$z_{ij} = \frac{1}{c_{ij}} \sum_{\forall x} c_{ij} X \quad (85.12)$$

- Step 11 If one among the termination criteria is met then go to Step 12. Else, go to Step 3.
- Step 12 Give the particle output with the minimum fitness value in the final generation.

85.3.4 Modified Artificial Neural Network (ANN)

Modified Artificial Neural Network (ANN) is used for the purpose of classification in the proposed methodology. Optimization task involves the training of an Artificial Neural Network (ANN) where the result provides bias set and optimal weight of the network. Extensive use of modified Levenberg-Marquardt (LM) algorithm is aimed at training the Artificial Neural Networks (ANN). It comprises of three layers, namely input layer, hidden layer, and output layer. For finding the solution, each individual neuron works in parallel.

Here x_1, x_2, \dots, x_n refer to the n inputs to the artificial neuron and w_1, w_2, \dots, w_n are the weights contained in the input links as illustrated in Fig. 85.1. The total input I obtained by the summation of the artificial neuron is computed as

$$I = w_1x_1 + w_2x_2 + \dots + w_nx_n \tag{85.13}$$

$$I = \sum_{i=1}^n w_i x_i \tag{85.14}$$

The sum is sent on to a nonlinear filter \emptyset known as Activation function, or Transfer function, or Squash function that provides the output. The output Y is expressed as:

$$Y = \emptyset(I) \tag{85.15}$$

In this research work, modified Levenberg-Marquardt (LM) gradient descent algorithm is employed widely for training the Artificial Neural Networks (ANN).

The performance index $F(w)$ to be reduced is defined as the summation of squared errors between the target outputs and the simulated outputs of the network, such as:

$$F(w) = e^T e \tag{85.16}$$

Where $w = [w_1, w_2, \dots, w_N]$ contains all the weights of the network, e refers to the error vector consisting of the error for every training example.

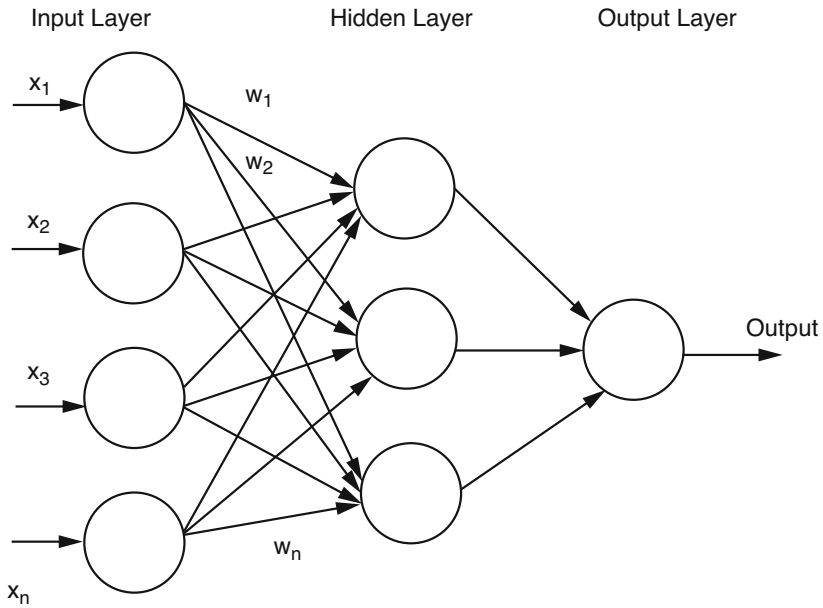


Fig. 85.1 Basic Structure of artificial neural network

Taking performance index to be $F(w)$ employing the Newton method, it is expressed as:

$$w_{k+1} = w_k - A_k^{-1} \cdot g_k \tag{85.17}$$

$$A_k = \nabla^2 F(W)|_{w=w_k} \tag{85.18}$$

$$g_k = \nabla F(W)|_{w=w_k} \tag{85.19}$$

$$[VF(W)]_j = \frac{\partial F(w)}{\partial w_j} \tag{85.20}$$

$$= 2 \sum_{i=1}^N e_i(W) \cdot \frac{\partial e_i(w)}{\partial w_j} \tag{85.21}$$

The gradient can be written as

$$\nabla F(x) = 2J^T e(w) \tag{85.22}$$

where

$$J(w) = \begin{bmatrix} \frac{\partial e_{11}}{\partial w_1} & \frac{\partial e_{11}}{\partial w_2} & \dots & \frac{\partial e_{11}}{\partial w_N} \\ \frac{\partial e_{21}}{\partial w_1} & \frac{\partial e_{21}}{\partial w_2} & \dots & \frac{\partial e_{21}}{\partial w_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_{kp}}{\partial w_1} & \frac{\partial e_{kp}}{\partial w_2} & \dots & \frac{\partial e_{kp}}{\partial w_N} \end{bmatrix} \tag{85.23}$$

$J(w)$ is known as the Jacobian matrix.

Next, the Hessian matrix is to be found. The k, j elements of the Hessian matrix are given as:

$$[\nabla^2 F(w)]_{k,j} = \frac{\partial^2 F(w)}{\partial w_k \partial w_j} \tag{85.24}$$

The Hessian matrix can then be represented as below:

$$\nabla^2 F(w) = 2J^T(W).J(W) + S(W) \tag{85.25}$$

Where

$$S(w) = \sum_{i=1}^N e_i(w). \nabla^2 e_i(w) \tag{85.26}$$

If it is assumed that $S(w)$ is small, the Hessian matrix can be approximated as:

$$\nabla^2 F(w) \approx 2J^T(w)J(w) \tag{85.27}$$

Making use of (27) and (18), the Gauss-Newton technique is obtained as:

$$W_{k+1} = W_k - [J^T(w_k).J(w_k)]^{-1} J^T(w_k) e(w_k) \tag{85.28}$$

The benefit of Gauss-Newton technique is that it does not need the computation of second derivatives.

An issue in the Gauss-Newton technique is that the matrix $H = JTJ$ might not be reversible. This can be resolved by making use of the modification below. The Hessian matrix can be expressed as:

$$G = H + \mu I \tag{85.29}$$

Assume that the eigenvalues and eigenvectors of H include $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ and $\{z_1, z_2, \dots, z_n\}$. Then:

$$Gz_i = (\lambda_i + \mu)Z_i \quad (85.30)$$

Hence, the eigenvectors of G are similar to the eigenvectors of H , and the eigenvalues of G are $(\lambda_i + \mu)$. The matrix G is positive definite by getting μ increased until $(\lambda_i + \mu) > 0$ for all i and hence the matrix will become invertible. This results in the Levenberg-Marquardt algorithm

$$w_{k+1} = w_k - [J^T(w_k)J(w_k) + \mu I]^{-1} J^T(w_k) e(w_k) \quad (85.31)$$

$$\Delta w_k = [J^T(w_k)J(w_k) + \mu I]^{-1} J^T(w_k) e(w_k) \quad (85.32)$$

As it is known, the learning parameter, μ , becomes the illustrator of steps of the original output movement to necessary output. In the standard LM technique, μ refers to a constant number. This article makes a modification to the LM technique with μ as:

$$\mu = 0.01e^T e \quad (85.33)$$

Where e refers to a $k \times 1$ matrix and hence $e^T e$ is a 1×1 and so $[J^T J + \mu I]$ becomes invertible. Hence, if the original output is far compared to the necessary output or in a similar manner, errors are large in number, it converges to the required output with big steps. Similarly, if the measurement of error is less then the original output nears the required output with soft steps. This way, the error oscillation is much decreased.

Algorithm

Step 1: Initialize ANN Weights.

Step 2: Calculate the sum of the squared errors over all inputs $F(w)$

Step 3: Solve (33) to get the increment of weights Δw

Step 4: Recompute the sum of squared errors $F(w)$

Step 5: Utilizing $w + \Delta w_k$ to be the trial w , and decide

Step 6: IF trial $F(w) < F(w)$ in Step 2 THEN

$w = w + \Delta w_k$.

Step 7: Go back to Step 2

Step 8: ELSE

Step 9: Go back to Step 4 Step 10: END IF

85.4 Experimental Results

The Matlab simulation environment is utilized for implementing the newly introduced research approach. The performance metrics that are taken into consideration are accuracy, recall, precision, f-measure, and time complexity. The comparison is done between the newly introduced EPSOGA-MANN scheme and the available techniques such as AABC-ITSVM, APSO-SVM, and APSO-NAIVES.

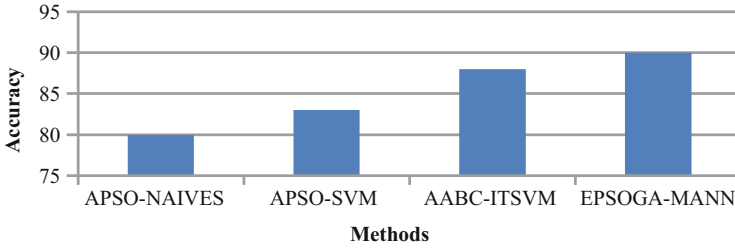


Fig. 85.2 Accuracy

Accuracy

It is defined to be the degree of correct detection. That is less false positive rate. The accuracy is calculated as follows:

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \tag{85.34}$$

As illustrated in Fig. 85.2 above, it can be noticed that the comparison metric is assessed employing the available and newly introduced techniques in terms of accuracy. The newly introduced Enhanced Particle Swarm Optimization with Genetic Algorithm (EPSOGA) chooses the best features. These features are used in modified ANN training and testing stage to generate results with more relevance. The result shows that the newly introduced system achieves superior classification results with EPSOGA.

Precision

Precision is described as the ratio of the true positives to both true positives and false positives results for imposition and actual features. It is expressed as below:

$$\text{Precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|} \tag{85.35}$$

As illustrated in Fig. 85.3, it can be seen that the comparison metric evaluation is done on the already available and newly introduced technique in terms of precision. The result shows that the newly introduced system achieves superior classification results with EPSOGA–MANN algorithm.

Recall

Recall value is calculated on the root of the data retrieval at true positive forecast, false negative. Generally, it can be computed as

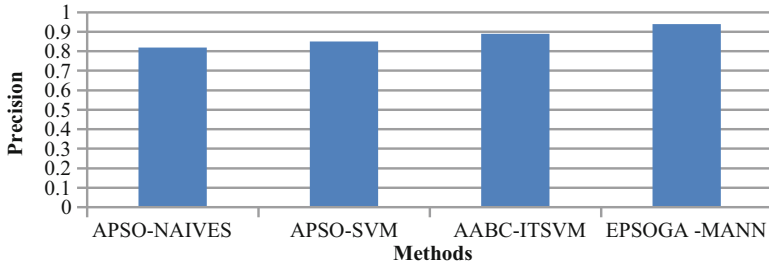


Fig. 85.3 Precision

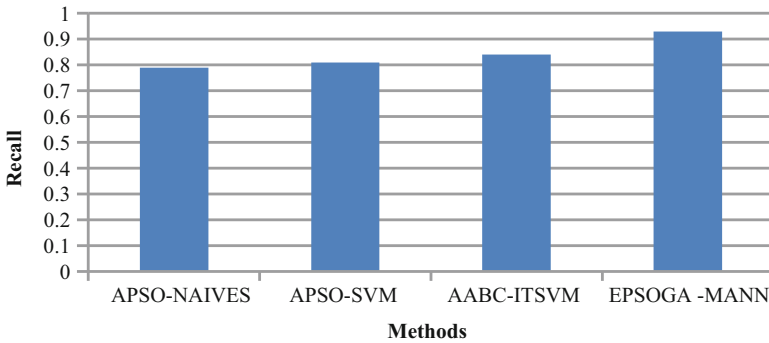


Fig. 85.4 Recall

$$\text{Recall} = \frac{T_P}{T_{P+F_N}} \tag{85.36}$$

As seen from Fig. 85.4 above, it can be noticed that the assessment of the call metric comparison on the already available and newly introduced technique is performed. The result shows that the newly introduced system achieves greater classification results with EPSOGA–MANN algorithm.

F-Measure

It provides the measure of the accuracy of a test. It takes both the precision *p* and the recall *r* of the test into consideration for the score computation.

As shown in Fig. 85.5 above, it can be noticed that the comparison metric is assessed on the available and newly introduced technique in terms of f-measure. The result shows that the newly introduced system achieves a greater f-measure result with EPSOGA–MANN algorithm.

Time Complexity

The system is called better when the algorithm gives minimum time complexity.

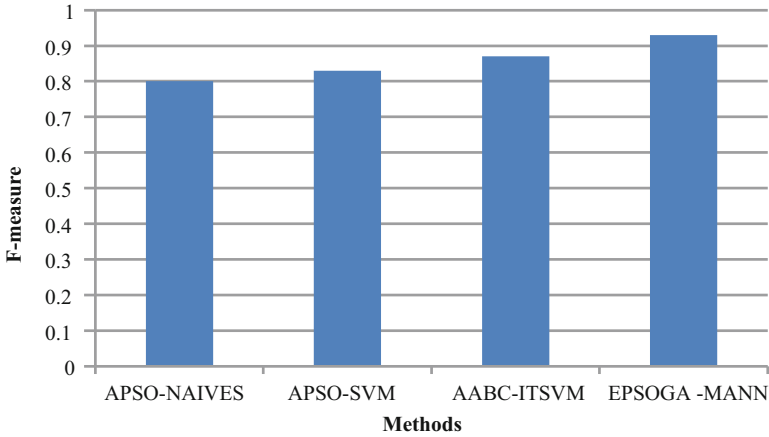


Fig. 85.5 F-measure

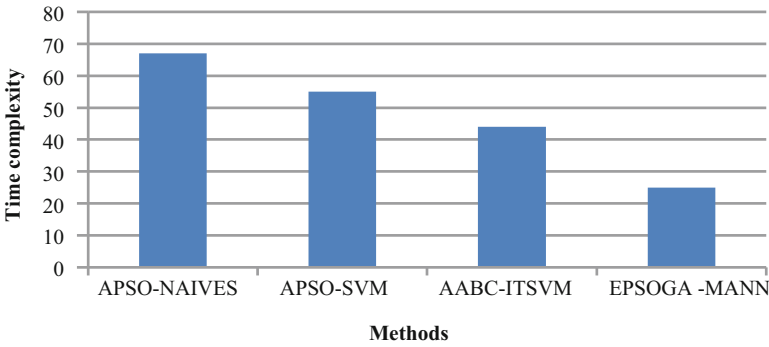


Fig. 85.6 Time complexity

As illustrated in Fig. 85.6, the graph shows the features chosen are assessed on the particular dataset. This way, the result shows that the newly introduced EPSOGA-MANN technique offers superior performance compared with the already available algorithms.

85.5 Conclusion

In big data analytics, the high dimensionality and the streaming behavior of the inward data increase the computational issues faced in data mining. Big data may not just be massive in volume but their structuring is done in several columns leading to high dimensionality in feature space. To solve this problem, the proposed system designed an Enhanced Particle Swarm Optimization with Genetic Algorithm-

Modified Artificial Neural Network (EPSOGA–MANN) approach. The newly introduced system comprises of three stages, which include preprocessing, feature selection, and classification. In the first phase, Fuzzy C Means (FCM) clustering algorithm is used to handle the noise data efficiently and discard the irrelevant features. Then, the optimal features are selected by using EPSOGA optimization algorithm. Depending on the features chosen, the classification is performed using the Modified Artificial Neural Network (ANN) for the given large volume of dataset. The experimental result demonstrates that the newly introduced system yields excellent performance in terms of greater accuracy, recall, precision, f-measure, and lesser time complexity with EPSOGA–MANN technique.

References

1. Gaber MM, Zaslavsky A, Krishnaswamy S (Jun. 2005) Mining data streams: a review. *ACM SIGMOD Rec* 34(2):18–26
2. Fan W, Bifet A (Dec. 2012) Mining big data: current status, and forecast to the future. *SIGKDD Explor* 14(2):1–5
3. Zliobaite I, Bifet A, Pfahringer B, Holmes G (2011) Active learning with evolving streaming data. *Proc Eur Conf Mach Learn Knowl Discov Databases* 3:597–612
4. Somol P, Pudil P (2000) Oscillating search algorithms for feature selection. In: *ICPR 2000*, vol 02. IEEE Computer Society, Los Alamitos, pp 406–409
5. Simpson R, Abakarov A, Teixeira A (2008) Variable retort temperature optimization using adaptive random search techniques. *Food Control* 19(11):1023–1032
6. Das S (2001) Filters, wrappers and a boosting-based hybrid for feature selection. In: *ICML '01: Proceedings of the eighteenth international conference on machine learning*, pp 74–81
7. Wills AS (2000) A new protocol for the determination of magnetic structures using simulated annealing and representational analysis (SARAh). *Phys B Condens Matter* 276:680–681
8. Cordeau JF, Laporte G (2003) A tabu search heuristic for the static multi-vehicle dial-a-ride problem. *Transp Res B Methodol* 37(6):579–594
9. Bakirtzis AG, Biskas PN, Zoumas CE, Petridis V (2002) Optimal power flow by enhanced genetic algorithm. *IEEE Trans Power Syst* 17(2):229–236
10. Dorigo M, Birattari M (2011) Ant colony optimization. In: *Encyclopedia of machine learning*. Springer, Boston, pp 36–39
11. Yaseen SG, Al-Slamy NM (2008) Ant colony optimization. *IJCSNS* 8(6):351
12. Anh Thu Vu (2014) Distributed adaptive model rules for mining big data streams. *Big Data (Big Data)*, IEEE, pp 345–353
13. Peralta D, del Río S, Ramírez-Gallego S, Triguero I, Benitez JM, Herrera F (2015) Evolutionary feature selection for big data classification: a mapreduce approach. *Math Probl Eng* 2015
14. Xin J, Wang Z, Qu L, Wang G (2015) Elastic extreme learning machine for big data classification. *Neurocomputing* 149:464–471
15. Fong S et al (2014) A scalable data stream mining methodology: stream-based holistic analytics and reasoning in parallel. *Computational and business intelligence (ISCBI)*, 2014 2nd international symposium on, IEEE
16. Fong S, Wong R, Vasilakos AV (2016) Accelerated PSO swarm search feature selection for data stream mining big data. *IEEE Trans Serv Comput* 9(1):33–45
17. Izakian H, Abraham A (2011) Fuzzy C-means and fuzzy swarm for fuzzy clustering problem. *Expert Syst Appl* 38(3):1835–1838

Chapter 86

Feature Selection Techniques for Email Spam Classification: A Survey



V. Sri Vinitha and D. Karthika Renuka

Abstract In this digital world, most of the communication is done only through the Internet. Email is widely used for exchanging information not only for personal communication but also has an important part in business communication because of its effectiveness, fastness, and cost-effective mode of communication. Spam email is the serious problem on the Internet; when users click on to the spam mail, it starts spreading viruses in the user system, consumes lot of network bandwidth and email storage space, and steals user's confidential data. Feature selection approach selects the best features from the dataset which removes irrelevant, redundant, and noisy data. The proposed paper offers email spam detection which incorporates various feature selection approaches like Information Gain, Correlation-Based Feature Selection, Genetic Algorithm, Ant Colony Optimization, Artificial Bee Colony, Particle Swarm Optimization, Cuckoo Search Algorithm, Harmony Search Algorithm, etc.; when classification is done after feature selection, it will enhance the performance of spam filtering.

Keywords Feature selection · Information gain · Genetic algorithm · Artificial bee colony · Ant colony optimization · Particle swarm optimization · Cuckoo search algorithm · Harmony search algorithm

Abbreviations

ABC	Artificial Bee Colony Optimization
BoW	Bag-of-Word
CGA	Compact Genetic Algorithm
EA	Evolutionary Algorithm

V. S. Vinitha (✉)

Bannari Amman Institute of Technology, Sathyamangalam, Erode, India

D. K. Renuka

PSG College of Technology, Coimbatore, Tamil Nadu, India

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,

https://doi.org/10.1007/978-3-030-24051-6_86

925

F-GSO	Firefly-Group Search Optimizer
HKSVM	Hybrid Kernel based Support Vector Machine
HSA	Harmony Search Algorithm
KNN	K-Nearest Neighbors
LR	Logistic Regression
MLP	Multi-Layer Perceptron
MLP-NN	Multi-Layer Perceptron Neural Network
NB	Naïve Bayes
PCA	Principal Component Analysis
PNN	Probabilistic Neural Network
PoS	Part-of-Speech
PSO	Particle Swarm Optimization
PU	Positive Unlabeled
RF	Random Forests
SCS	Stepsize Cuckoo Search
SMO	Sequential Minimal Optimization
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
TREC	Text Retrieval Conference
UCI	UC Irvine

86.1 Introduction

Feature selection [1] is a technique for discovering a minimal number of features f from the original F features of an email spam dataset. Some features are relevant for spam detection, but some are not with repetition and also it can be continuous, discrete, or nominal. It is recognizable to determine the finest attributes from the new email dataset that comprises the features with smallest number of dimensions which contributes by removing irrelevant data, reducing dimensionality to increase the accuracy and to improve the performance. This technique has optimal number of features which attain same or better results.

The features that are related to the spam detection and impact on spam detection are known as relevant features, while the rest cannot perform their role. If the features do not have any impact on spam detection, they are irrelevant features. If there are features with repetition, then they are redundant and even the features may be different. The real-world dataset used for spam detection may contain noise, irrelevant or ambiguous features where feature selection plays a vital role. Many algorithms, approaches, and methods are available for feature selection.

Feature selection algorithms [2] are predominantly characterized into three methods. They are:

- Filter method
- Wrapper method
- Embedded method.

86.1.1 Filter Method

Filter method is to be applied for choosing the significant features that have to be done before classification. It is used to determine the best features from the input spam email. It is independent of any classification algorithm and filters data based on the selective criteria. The input of the filter is the attributes of email dataset. Based on the scores obtained from various statistical tests, features which are significant in determining the outcome variable can be selected. If there are more number of features, then the filter method can be more suitable because of the high computational efficiency. Some of the filter methods are correlation-based methods, mutual information-based methods, information gain, chi-squared test, etc.

86.1.2 Wrapper Method

Wrapper method is suitable for dataset that contains less amount of attributes. For the given email dataset, it finds the suitable attributes and trains a model using them. Based on the dependencies among features, attributes can be added or removed from the subset. It provides better result when compared with the filter method. But it requires more computational resources than the filter method and more appropriate for small training datasets. Few of them are Sequential Forward Selection, Genetic Algorithms, Stepwise Regression, Backward Elimination Method, etc.

86.1.3 Embedded Method

Embedded method has been projected to incorporate the benefits of filter method and wrapper method. In this model, some good features will be selected from an email spam dataset by using the filter method. Then wrapper method is applied on those selected features to acquire the best feature. For feature selection, one of two methods, such as the subset selection or the feature ranking method, can be used. The set of possible features is selected based on the criterion, forms the optimal subset in the subset selection method, and ranks the features according to the criterion in the feature ranking process. Weighted Naïve Bayes, Sequential Forward Selection, Artificial Neural Networks, etc., come under the embedded method.

86.2 Overview of Email Spam Detection

In recent years [3], email has become a platform that is extensively used in the Internet for communication. It is an electronic messaging system used to transfer message from one user to another. In the email, spam is the major concern, which

Table 86.1 Detailed analysis of datasets used in spam email classification

Dataset	Dataset sample Size
PU	Total 7101 email (spam = 3020 and ham = 4081)
Custom	It varies from study to study
SpamBase	Total 4601 emails (spam = 1813 and ham = 2788)
Enron spam Corpus	Total 30,041 emails (spam = 13,496 and ham = 16,545)
SpamAssasin	Total 10,744 emails (spam = 3793 and ham = 6951)
TREC	Total 92,189 emails (spam = 52,790 spam and ham = 39,399)
CCERT	Total 34,360 emails (spam = 25,088 and ham = 9272)
LingSpam	Total 3252 emails (spam = 841 and ham = 2412)

transmits messages to bulk amount of beneficiaries. Spam email is also called as junk mail. Spammers usually collect these addresses from websites for spreading malware and sending phishing emails for stealing user confidential data. It consumes email storage space and wastes user time in opening and deleting the junk emails.

86.2.1 Dataset

There are various datasets available in the UCI Repository to detect spam mails. The dataset contains spam and ham mails. Some of them [4] are listed in the table (Table 86.1).

86.2.2 Evolutionary Algorithms

Evolutionary algorithm uses nature-inspired approach for optimization. It follows the behavior of living organisms to solve the problem and is inspired by the concepts in Darwinian Evolution and modern genetics. Evolutionary algorithm is intended for resolving a problem more quickly which will consume more time for thorough processing. There are four inclusive steps in EA, which are Selection, Mutation, Crossover, and Accepting. In EA, appropriate members will persist and increase, while irrelevant members will perish and not contribute for further generations. From the given population of individuals, natural selection is made by environmental pressure to rise the appropriateness of the population. Fitness measure is applied to the appropriate solutions which were randomly created. With this measure, the better possible solutions are applied with recombination or mutation and given to the next generation. Recombination operator is enforced on parents and results in the children. Mutation operator is enforced on one candidate to produce another new candidate. Depending on their fitness measure in the next generation, recombination and mutation replace old ones with a new candidate. This process will be iterated until the best solution is found.

86.2.3 Classification Algorithms

After selecting the finest attributes from an email spam dataset, Machine Learning algorithms are used for performing classification. Various algorithms are available to classify the non-spam and spam emails. Classification algorithms include Decision Tree, AdaBoostJ48, Naïve Bayes (NB), Support Vector Machine (SVM), Random Forests (RF), Neural Networks, and Multi-Layer Perceptron (MLP). Then the performance is evaluated by using different metrics such as accuracy, sensitivity, and specificity.

86.3 Related Works

Some of the research works done on predicting the spam email by employing various Feature Selection techniques to select the best features for performing classification are deliberated in the following section:

Email is the instantaneous method for exchanging information through the Internet. Spam mails contain mischievous code to steal personal information about the user and also to infect user's system through spreading viruses. In order to reduce the consequences of spam mail, Vrinda Sharma [5] proposed a Term Frequency and Inverse Document Frequency (TF-IDF) and Information Gain for efficient feature selection. Then, the result of these two feature selection is applied on four classification algorithms, namely, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors (KNN), and Random Forest. It is tested on different datasets such as DBWorld E-Mails, LingSpam, and Enron dataset using classification algorithms like Naïve Bayes, KNN, Random Forests, and SVM. Random Forests and SVM provide a better result, but SVM takes more time. But Naïve Bayes and KNN are improved in terms of accuracy and time.

Spam is a major concern on the today's Internet. To classify the spam emails, four feature selection techniques and Machine Learning algorithms are used for classification. Reshma Varghese [6] recommended Bag-of-Word (BoW)s, Bigram Bag-of-Word (BoW)s, Part-of-Speech (PoS) Tag, and Bigram PoS Tag for extracting the features. The Naïve Bayes score is used to eliminate the rare features. Enron dataset is taken as the input. Features are selected by Information Gain and form matrix using Term Frequency – Inverse Document Frequency (TF-IDF). For classification, AdaBoostJ48, Random Forest, and Popular Linear Support Vector Machine (SVM), called Sequential Minimal Optimization (SMO), are used and yield an accuracy of 0.932, 0.911, and 0.750 for Adaboost, Random Forest, and SMO. Adaboost provides good results with ensemble model.

Email is one of the fastest modes of communication used on a daily basis by millions of people. However, the number of email users has increased resulting in dramatic increase in spam mails over the past few years. P.U. Anitha [7] proposed an efficient spam classification technique using Naïve Bayes classifier and Compact

Genetic Algorithm (CGA) by using SpamBase and LingSpam datasets. It contains training and testing phases. During the training phase, best features are selected using hybrid Cuckoo Search and Genetic Algorithm. After selecting the best features, classification is done by using Naïve Bayes algorithm. Performance was compared with existing techniques like Particle Swarm Optimization (PSO). The comparison indicates that the proposed system using hybrid optimization provides better accuracy.

Email is one of the important ways of exchanging information. Spam is serious concern in today's Internet. So there is a need to filter the spam emails. Issam Dagher [7] recommended spam filtering using Kernel Principal Component Analysis. It is implemented using a Public Corpus extracted from the University of California-Irvine Machine Learning Repository. The best features are extracted using PCA. For classification using Support Vector Machine and Naïve Bayes, different training and testing sets are used. The spam mails are correctly classified for more number of trials and it takes less time comparable to PCA. Kernel PCA provides the best performance in terms of accuracy. The accuracy of the Bayes detector was high, but it takes more time for classifying large number of features.

Spam is the major problem faced by most of the email users, as it consumes large amount of email storage and steals all users' personal data. Therefore, a filter is needed to block these spam emails. In the dataset, not all the features are relevant for spam classification. Thus suitable features should be extracted for further processing. Mehdi Zekriyapanah Gashti [8] chose various datasets such as SpamBase, LingBase, and PU1 and applied the Harmony Search Algorithm (HSA) to select the best features. Selected features help to improve the accuracy of its predictions. Then, Decision Tree is used for classifying the selected features. The proposed model on SpamBase dataset provides an accuracy of about 95.25% which is better than SVM, J48, MLP, and NB. And also, the accuracy of proposed model on LingSpam and PU1 dataset provides better result than LR, NB, and SVM.

Email has established a significant role in exchanging information because of its fastest and cost-effective way of communication. It plays a vital role in both personal and business communication. The rapid growth of email has generated several issues. From past decades, spam emails start spreading tremendously. These spam emails spread malware in user system and steal personal data by sending phishing emails. So, there is a need for efficient filter to classify the spam and ham mails. Harjot Kaur [9] proposes MLP for classifying spam emails. MLP takes more execution time and degrades the performance of algorithm. So in future work, refined MLP along with N-Gram-based feature selection is used to remove noise and outliers in the dataset and for selecting the best features from the corpus.

The communication tool is attacked by intruders for sending unwanted spam emails. Several spam filtering techniques exist, but still the problem survives. Masoome Esmaeili [10] addresses this issue by implementing the Bayesian method and PCA to filter these spam emails from the user inbox. Forty spam and fifty non-spam emails are considered in the training step and extracted the features and saved them with their frequencies in a local dictionary. Then, they were classified by

using the Bayesian method and compared its result with various feature selection techniques. The ratio method was applied on the original dictionary in the preprocessing step to remove the irrelevant features. Then GA) was applied on modified dictionary and obtained 97.76% with 3400 features.

In this digital world, spam causes serious problem to the Internet users. In this paper, T. Kumaresan [11] suggested a modified Cuckoo Search called Stepsize Cuckoo Search (SCS) and Support Vector Machine for spam email classification. SCS is used which not only speeds up the convergence of the algorithm but also allows us to find the optimal features from the SpamBase dataset. Then the classification is done by using the Support Vector Machine. For the effectiveness of classification, three different kernels, such as linear, polynomial, and quadratic, are used. The proposed system is evaluated by different metrics like precision, recall, and accuracy, and the result shows that it provides better result when compared with the existing classification technique.

Due to cost-effective communication, email is widely used for personal and business communication for transferring messages. Spam has become a major problem because it causes unnecessary traffic and security threats. Several techniques have been deployed to block these spam emails. Shashi Kant Rathore [12] proposes the hybrid Bayesian algorithm and swarm intelligence for recognizing spam mails. Best features from the LingSpam dataset are selected by using swarm intelligence and the classification is done by using the Naïve Bayes algorithm. This approach takes static values of probabilities for each token. So, an automated trained filter can also be maintained by including Nature-based optimization techniques such as Artificial Bee Colony and Spider Monkey Optimization. From this, the best tokens can be classified to recognize the spam mails.

In this e-world, email stands out for communication in the Internet. Because of its popularity, it is misused by people for sending unwanted messages to large number of recipients. These emails are called as spam emails. Spam email lessens the productivity, consumes extra storage in the mailbox, takes up a lot of time for opening and deleting the mails, spreads viruses, and steals user's data through phishing emails. So, there is a need to block the spam mails from entering the user's inbox. Shradhanjali [13] suggested a novel method using Support Vector Machine and feature extraction. The proposed system obtains an accuracy of about 98% with the test datasets.

Spam email causes severe problem to the Internet community that threatens network bandwidth and productivity of the users. T. Kumaresan [14] recommended a framework using S-Cuckoo and Hybrid Kernel based Support Vector Machine (HKSVM). At first, textual features are selected from the LingSpam dataset using Term Frequency and for images, correlogram and wavelet moment are taken. The optimal features are selected using hybrid S-Cuckoo Search. After selecting the features, classification is done using HKSVM. Then, the performance is analyzed by using evaluation metrics such as precision, recall, and accuracy. Experimental result shows that the proposed HKSVM provides better result when compared to other SVM-based models (Table 86.2).

Table 86.2 Performance analysis of email spam detection using feature selection and classification techniques

Author	Feature selection used	Classification techniques used	Dataset	Evaluation metrics
Vrinda Sharma et al. [5]	TF-IDF and information gain	Naïve Bayes, SVM, KNN, and Random Forests	DBWorld E-mails, LingSpam and Enron6 datasets	Accuracy: NB – 0.8524, KNN – 0.8196, RF – 0.9852, SVM – 1
Reshma Varghese et al. [6]	Information gain	Adaboostj48, random forest, and sequential minimal optimization (SMO)	Enron spam dataset	Accuracy: Adaboost – 0.932, Random Forest – 0.911, SMO – 0.750
Mehdi Zekriyapanah Gashti [8]	Harmony search algorithm	Decision tree	SpamBase, LingSpam, and PUI datasets	Accuracy: SpamBase – 0.9525, LingSpam – 0.9980, PUI dataset – 0.9712
Harjot Kaur et al. [9]	N-gram-based feature selection	Multi-layer perceptron neural network (MLP-NN) and Support Vector Machine (SVM)	Enron dataset	Accuracy: SVM – 0.6466, MLP – 0.7809
Shradhanjali et al. [13]	Features that matched the word from dictionary are extracted and are mapped using vocab file	Support vector machine	Apache public corpus	Accuracy – 0.98
Dhanaraj Karthika Renuka et al. [15]	F-GSO algorithm	Decision tree rule, Naïve Bayes, and neural network	SpamBase dataset	Accuracy – 0.9883
Masooma Esmaili et al. [10]	Principal component analysis	Naïve Bayes	Forty spam and fifty non-spam emails	Accuracy – 0.9687

(continued)

Table 86.2 (continued)

Author	Feature selection used	Classification techniques used	Dataset	Evaluation metrics
Issam Dagher et al. [7]	(PCA)	Support vector machine, Naïve Bayes	University of California-Irvine Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/SMS+Spam+Collection)	Accuracy: SVM – 0.9621, NB – 9695
T. Kumaresan et al. [14]	Stepsize-Cuckoo Search (SCS)	SVM	LingSpam Dataset	Accuracy – 0.96
T. Kumaresan et al. [14]	S-Cuckoo Search	Hybrid kernel based support vector machine (HKSVM)	LingSpam dataset	Accuracy – 0.97235
Shashi Kant Rathore et al. [12]	ABC (Artificial Bee Colony Optimization), SMO (Spider Monkey Optimization)	Naïve Bayes	LingSpam dataset	Accuracy – 0.92
Mohammad Zavvar et al. [16]	Hybrid particle swarm optimization algorithms and artificial neural network	Support vector machine	SpamBase dataset	Accuracy – 0.8742
D.Karthika Renuka et al. [17]	Ant Colony optimization	Naïve Bayes	SpamBase dataset	Accuracy – 0.84
Masurah Mohamad et al. [18]	Term frequency inverse document Frequency (TF-IDF)	Rough set theory	169 emails comprising of texts and images	Accuracy – 0.848
S. Kumar et al. [19]	Particle swarm optimization (PSO)	Probabilistic neural Network (PNN)	90 emails	Accuracy – 0.90
Sorayya Mirzapour Kalaibar et al. [20]	Genetic algorithm	Bayesian network and KNN classifiers	SpamBase dataset	Accuracy: Bayesian network – 0.935, KNN – 0.886

86.4 Conclusion

This paper provides an overview of various feature selection techniques that can be used for email spam detection. In the input email dataset, all the attributes are not relevant for detecting the spam mails. Some features are relevant, but some are irrelevant. So, there is a need to eliminate the irrelevant features that lessen the

execution time and provide better accuracy. Feature selection is used to select the best attributes from the spam email dataset. It reduces the dimension of the input and after that, it uses classification techniques to classify the spam emails, which helps in improving the performance of spam detection.

Acknowledgments Our sincere thanks to the University Grants Commission (UGC), Hyderabad, for granting the funds to carry out this work.

References

1. Xue B, Zhang M, Browne WN, Yao X (2016) A survey on evolutionary computation approaches to feature selection. *IEEE Trans Evol Comput* 20(4):606–626
2. Jain D, Singh V (2018) Feature selection and classification systems for chronic disease prediction: a review. *Egypt Inform J Elsevier* 19(3):179–189
3. Bhuiyan H, Ashiquzzaman A, Juthi TI, Biswas S, Ara J (2018) A survey of existing E-mail spam filtering methods considering machine learning techniques. *Global J Comp Sci Technol C Softw Data Eng* 18(2):21–29
4. Mujtaba G, Shuib L, Raj RG, Majeed N, Al-Garadi MA (2017) Email classification research trends: review and open issues, *IEEE Access*, pp 9044–9064
5. Sharma V, Poriye M, Kumar V (2017) Various classifiers with optimal feature selection for email spam filtering. *Int J Comput Sci Commun* 8(2):18–22
6. Varghese R, Dhanya KA (2017) Efficient feature set for spam email filtering. *IEEE 7th international advance computing conference*, pp 732–737
7. Dagher I, Antoun R (2017) Ham – Spam Filtering using Kernel PCA. *Int J Comput Commun* 11:38–44
8. Mehdi Zekriyapanah Gashti (2017) FHSA. *Eng Technol Appl Sci Res* 7(3):1713–1718
9. Kaur H, Prince Verma E (2017) E-mail spam detection using refined MLP with feature selection. *Int J Modern Educ Comput Sci* 9:42–52
10. Esmaili M, Arjomandzadeh A, Shams R, Zahedi M (2017) An anti-spam system using naive Bayes method and feature selection methods. *Int J Comput Appl* 165(4):1–5
11. Kumaresan T, Palanisamy C (2017) E-mail spam classification using S-cuckoo search and support vector machine. *Int J Bio-Inspired Comput* 9(3):142–156
12. Rathore SK, Yada S (2017) A hybrid Bayesian approach with ABC to recognition of email SPAM. *Int J Comput Sci Mob Comput* 6(5):459–466
13. Shradhanjali, Verma T (2017) E-mail spam detection and classification using SVM and feature extraction. *Int J Adv Res Ideas Innov Technol* 3(3)
14. Kumaresan T, Saravanakumar S, Balamurugan R (2017) Visual and textual features based email spam classification using S-cuckoo search and hybrid kernel support vector machine. *Clust Comput* 22:33–46. Springer Publication
15. Renuka DK, Visalakshi P (2017) Weighted-based multiple classifier and F-GSO algorithm for email spam classification. *Int J Business Intelligence Data Mining* 12(3):274–298
16. Zavvar M, Rezaei M, Garavand S (2016) Email spam detection using combination of particle swarm optimization and artificial neural network and support vector machine. *Int J Modern Education Computer Science (IJMECS)* 8(7):68–74
17. Karthika Renuka D, Visalakshi P, Sankar T (2015) Improving E-mail spam classification using ant Colony optimization algorithm. *Int J Comput Appl* 22–26

18. Mohamad M, Selamat A (2015) An evaluation on the efficiency of hybrid feature selection in spam email classification. *IEEE international conference on computer, communication, and control. Technology* 227–231
19. Kumar S, Arumugam S (2015) A probabilistic neural network based classification of spam mails using particle swarm optimization feature selection. *Middle-East J Sci Res* 23(5):874–879
20. Kalaibar SM, Razavi SN (2014) Spam filtering by using genetic based feature selection. *Int J Comput Appl Technol Res* 3(12):839–843

Chapter 87

A Novel Paradigm Towards Exploration of Rechargeable WSN Through Deep Learning Architecture for Prolonging Network Lifetime



M. Ezhilarasi  and V. Krishnaveni 

Abstract Rapid development of energy efficient technique for wireless sensor network and its proliferation can relieve the energy constraints on sensor to a little extent but limited the lifetime of the batteries in the sensor. Upon exploration, rechargeable wireless sensor network has potential to mitigate this issue by prolonging the network through extraction of renewable energy to replenish the sensor on the deployed region. Energy-constrained deep learning architectures have to be utilized as optimization objective to the existing routing protocol to enhance the reliability and flexibility of the network. In this chapter, a novel deep learning algorithm, named as deep belief network, has been proposed to achieve energy efficiency. The dynamic source routing protocol is been employed on this paradigm; the mobile sink is utilized for data gathering and replenishing of energy in the cooperative manner towards data transmission. The deep belief network exploits the route with shortest path through information of the mobile sink on a random transmission. Priorization is employed to sensor nodes with least energy will be charged by the mobile node while computation of the node density. Furthermore, the proposed model eliminates the localization issues and latency issue of mobile node. Moving trajectory of the mobile sink is determined with optimal velocity control mechanism. The simulation results of the proposed architecture proves that the proposed paradigm exhibits a good performance in terms of throughput, latency, packet delivery ratio, routing overhead, and energy utilization of nodes compared with state of approaches.

Keywords WSN · Energy · Lifetime · Deep learning · Dynamic source routing

M. Ezhilarasi (✉)
Sri Ramakrishna Engineering College, Coimbatore, Tamil Nadu, India

V. Krishnaveni
PSG College of Technology, Coimbatore, Tamil Nadu, India
e-mail: vk.ece@psgtech.ac.in

© Springer Nature Switzerland AG 2020

L. Ashok Kumar et al. (eds.), *Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*,
https://doi.org/10.1007/978-3-030-24051-6_87

937

Abbreviation

WSN Wireless sensor network

87.1 Introduction

87.1.1 A Subsection Sample

Due to large exploration of wireless sensor networks (WSNs), energy and efficiency are considered to be important issues as they are powered by energy-limited battery capacity of the sensor nodes [1]. In recent years, many energy harvesting technique has been developed on random deployment characteristics and depending upon the number of nodes deployed with in the environment. In addition, routing protocol results in the data traffic towards the sink which deplete the battery and leads to energy hole problem [2]. Energy replenishing of the batteries in the sensor is carried out through exploration of rechargeable wireless sensor network [3]. Rechargeable wireless sensor network consists of number of node which can generate the energy in addition to battery through renewable sources. It reduces the impact of the energy harvesting protocol [4]. Mobile sink is used to carry the energy and data of the sensor nodes. In addition, deep learning architecture is used to determine the node, depleting faster due to data forwarding process and data storage process. It computes the large queue node to shift the data packets on effective path. Dynamic source routing protocol supports the identification of mobile sink trajectory with velocity estimation [5].

The deep learning architecture uses deep belief network model to exploit the route of least energy node and predicts the optimal velocity for node propagation in the path and prioritization [6]. The localization issue of mobile node is completely eliminated.

The remaining part of this chapter is sectioned as follows: The related work on energy efficient techniques of wireless sensor network is provided in Sect. 2. Architecture of the proposed framework utilizing deep learning model is given in Sect. 3. The simulation results of the entire work are presented in Sect. 4. Finally, Sect. 4 concludes the paper with summary and future research directions.

87.2 Related Works

In literature survey, many approaches related to the energy maximization and increasing the network lifetime in WSN has been analyzed in depth on several aspects is discussed below. The techniques are discussed below.

87.2.1 Distributed Cooperative Relaying Low Energy Adaptive Clustering Hierarchy Protocol

In this protocol, energy consumption is carried out based on cooperative relaying technique [7]. In this cluster head cooperates the other high energy node from the cluster to transmit the collected data to the sink. The protocol carries out this operation with less number of hops using multihop routing protocols. It fails to include the data gathering model and replenishing technique to heavy congested node as it depletes faster during transmission.

87.2.2 Bayesian Compressive Data Gathering Protocol

According to this protocol, the data compression is engaged to the data propagating in the network using Bayesian compression [8]. The path to the sink is also computed using chaos technique. It is a time varying protocol. It works on several constraints in a diversified manner on the center of deployment area. Multiple variable optimizations are done to avoid the inevitable error on the relay of the network with multiple time slots.

87.3 Proposed Model

In this section, a proposed paradigm is defined in detail through establishment of wireless sensor network on a random deployment model with composition of mobile sink. Mobile sink is used to gather data and replenish the energy to energy degraded nodes through utilization of the routing information. The architecture is as follows.

87.3.1 Network Model

Wireless sensor node is deployed randomly with arbitrary distance towards data transmission and reception. The nodes are bidirectional and it flows through either mesh or star topology. Each sensor node is employed with unique address using IPV6 addressing protocol. The node will be deployed with high degree of energy. During monitoring, data processing, and data transformation, the energy of the node depletes gradually. Each node will undergo state transmission from idle to transmitting and transmitting to receiving, and finally from receiving to sleep state. The nodes provide stronger coverage, and the information is stored with the help of base station; distance has high influence in bandwidth and energy constraints.

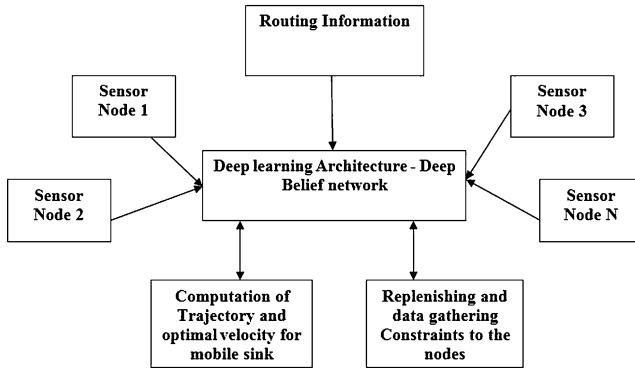


Fig. 87.1 Architecture of the proposed model

87.3.2 Mobile Sink

A mobile sink is modeled as both energy replenishing unit and data gathering unit. It collects data to transmit the base station from the sensed node. The mobile sink travels the path with either dynamic speed or optimal calculated speed on the specified trajectory.

Based on the received signal strength, the distance from signal source to receiver can be computed. The nodes are stored in the form of tier; the high tier nodes acquire' to store the fresh sink position and low tier nodes queries the high tier node whenever it requires. The mobile sink selects the path based on the tiers. In addition, it partitions the network into clusters in order to increase the reliability of the data transmission. The proposed model architecture is depicted in Fig. 87.1.

87.3.3 Dynamic Source Routing

The protocol determines the source route after accumulating all the information of the other nodes during route discovery phase [9]. It computes the path based on the data and queue length of each node. It protects the route reply storms and reduces the route error message. Cache information is shared with other nodes during node error which effectively reduces the energy depletion [10]. The energy depletion of the sink is calculated as follows

$$E_i = E_t (M_{-i} + 1) + E_r (K_{-i}) + E_i$$

The energy depletion equation helps in generating optimal routing and load balancing of the node against the various packet sizes.

87.3.4 Deep Belief Network

The deep belief network is employed to determine the several aspects such as density of node, energy depletion rate of the node, and node distance from the mobile sink. The architecture improves the reliability and flexibility by computing the energy-based probabilistic model in each layer of the architecture [11]. It includes velocity control scheme which depends on the acceleration constraints of a predefined trajectory arbitrarily shaped in a 2D space. The objective function to determine these constraints is as follows.

$$P(x;\theta) = \frac{f(x;\theta)}{\sum_x f(x;\theta)} = \frac{f(x;\theta)}{Z(\theta)}$$

$f(x;\theta)$ is a function to compute the distance of node and trajectory. $P(x;\theta)$ is a selected path to the highly energy depleted node.

The moving trajectories were constructed based on demand constraints trajectories which not only improve the throughput even they increase the data delivery latency for data gathering from single hop or multihop sensor node [12]. The trajectories are given by

$$P(A, B, C; \theta) = \frac{\exp(w_1 BC + w_2 AC)}{\sum_{A, B, C} \exp(w_1 BC + w_2 AC)}$$

All the sensor nodes transmit the data of mobile sink at most in any consecutive rounds if it has constraints of the expected path.

$$\text{Expected path} = \frac{\exp(w_1 BC) \exp(w_2 AC)}{Z(w_1, w_2)}$$

Most of the mobile sink charging scenarios, the difficulties are movement of the charger, time, and space. Traveling time of the charge is measured using parameter named as patrol cycle [13]. The optimal velocity is calculated as follows

$$P(x, h) = \frac{e^{-E(x, h)}}{\sum_{x, h} e^{-E(x, h)}}$$

The optimal velocity is determined in order to prolong the lifetime of sensor network by controlling the charging velocity of the mobile charger to the sensing nodes.

Input – E_n Energy of n th sensor node
 M_s – available charge in the Mobile Charger
 Output-Depletion rate of Sensor and Velocity rate to sensor
 Process
 Estimate the Patrolling time T
 Determine Time t between the nodes to measure charge level
 If (Energy level of node $n >$ threshold)
 Use optimal Charging velocity
 Else.
 Maximize the Charging velocity
 Calculate the depletion rate of the sensing node
 Depletion Difference = $D_i - D_{i-1}$
 If (Depletion difference $<$ threshold)
 Replace the Mobile charger
 Else
 Calculate the depletion rate of the sensor node.

Depletion rate is calculated in terms of linear correlation coefficient of the node, and it is given by

Linear correlation coefficient

$$Lc = \sum_{k=0}^n x^k a^{n-k}$$

where a is the node and k is the value of node density on specific features. Residual energy of the each path is calculated using.

$$E_{tx} = (\alpha_1 + \alpha_2 d^n) * r$$

where α_1 is the available buffer space, α_2 is the current packet load of the node, d^n is the distance between the nodes, and r is the radius of the node.

The sink node speed, which is traveling in the trajectory, is given by

$$S_v = \frac{1}{E} \sum_{t=1}^T \sqrt{(\alpha_1 + \alpha_2)^2} + \sqrt{L_c + (r)^n}$$

15 km/h speed used for set of simulations.

The process is repeated with determination of various neighbors at each hop and different radius to obtain best optimal value. It emphasizes the strength of the proposed routing and charging model of sink. The speed of the sink is not the main parameter affecting the network performance. The total number of nodes deployed in the network also affects the WSN performance as the network density

and total traffic loads are based on the network size [14, 15]. The nodes which have higher signal strength have more communication link and result in faster energy consumption.

87.4 Simulation Results

The proposed paradigm towards increasing network lifetime of WSN is carried out using deep learning architectures has been using NS2 Simulator in this section. On extensive analysis utilizing the performance metrics like throughput, packet delivery ratio (PDR), network overhead, and packet loss demonstrates the effective of the proposed architecture [16]. The simulation setup of the network is defined in Table 87.1.

The higher patrolling cycle provides maximum acceleration of mobile sink and increases the network lifetime. Performance evaluation on the proposed and existing model was performed and the results are tabulated in Table 87.2. The definition of the measure is as follows.

87.4.1 Throughput

It is defined as multiplication of transmission rate of the mobile sink and time taken for route request and route reply during the transmission in the single path. Throughput calculation informs the performance of the proposed charging mechanism.

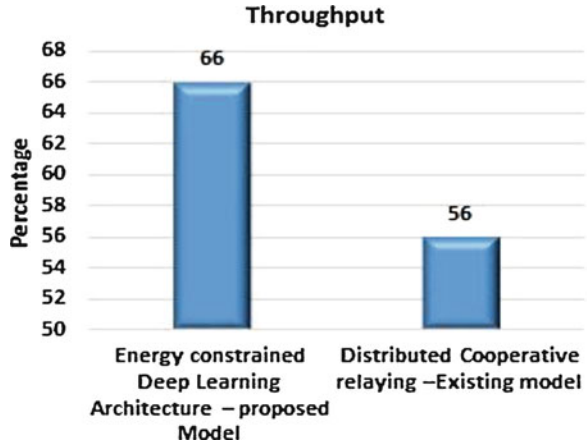
Table 87.1 Simulation setup

Simulation parameter	Value
Simulator	NS2
Topology size	750 m *750 m
Number of nodes	100
Bandwidth of the network	2.5Mbps
Network traffic type	CBR
Packet size	512 bytes
Simulation time	30 min

Table 87.2 Performance value of the proposed paradigm and existing models

Technique	Throughput	Overhead	Depletion rate	Energy utilization
	mbps	mbps	mbps	%
Energy-constrained deep learning architecture – proposed model	66	11.23	2	85
Distributed cooperative relaying – existing model	56	17.89	2.9	69

Fig. 87.2 Throughput of the proposed model against existing technique



87.4.2 Routing Overhead

It is defined in terms of the node energy difference between the transmitting nodes, sink node, and source node in the particular path for single packet transmission. Depletion rate formulation and energy conservation definition were stated in the algorithm of the proposed model.

The traffic of the packet load to the node generally increases as a function of the node’s proximity to the sink. Through deep learning approach, its communication structure is estimated. Hence, it helps better optimizing the performance of the network. The lifetime of the network gradually prolongs with the growing maximum acceleration in the charger and the patrol cycle through several layers of processing on the deep belief network.

According to the proposed work, the average energy consumption provides the projected duration of service of the wireless sensor network operation for lifetime. Figs. 87.2, 87.3, 87.4, and 87.5 explain the performance evaluation of the proposed model against existing model. The proposed deep learning framework of the data traffic provides the uniform and reduced energy consumption. The intensity of node failure minimizes the usage of mobile sink, as the sink continuously reaches the disconnected areas in the network. This way, it provides the most reliability and data delivery ratio.

87.5 Conclusion

The design and implementation of energy exploiting paradigm towards achieving optimal velocity on energy replenishing and data gathering of sensor information by mobile sink has been done. The proposed model utilized the deep learning architecture to determine the intermediate node which requires the energy through deep

Fig. 87.3 Overhead of the proposed model and existing technique

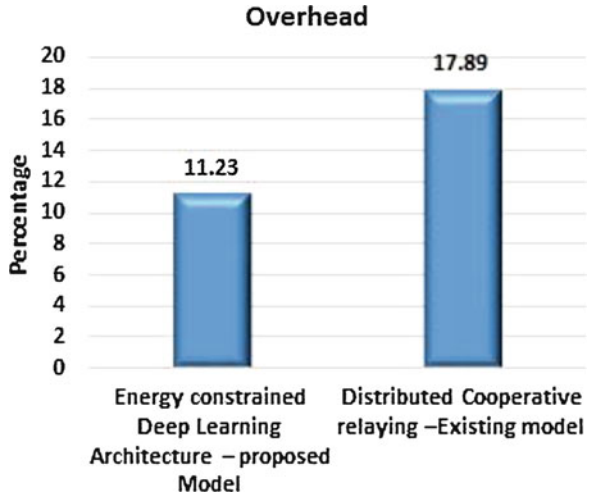
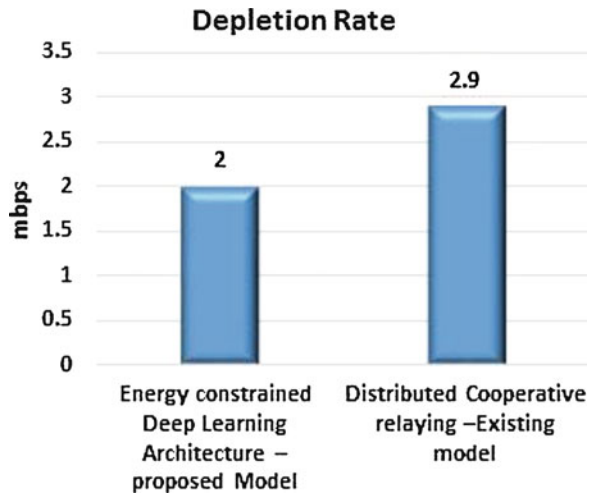
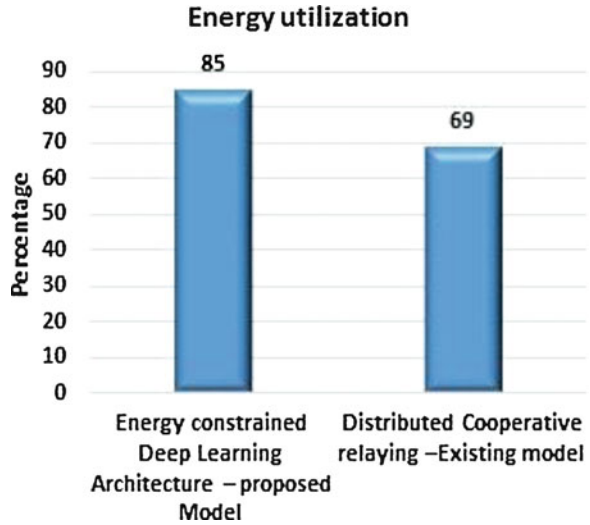


Fig. 87.4 Depletion of the proposed model against existing technique



belief network. The deep belief network computes the trajectory of the node. An extensive variety of various range with differing network sizes and sink speed are characterized and utilized. Simulation results prove that the proposed model outperforms the state-of-the-art approaches in all network evaluation metrics. In future, network has to be modeled to support multiple sink nodes.

Fig. 87.5 Energy utilization evaluation of the proposed model against existing technique



References

1. Duarte-Melo EJ, Liu M (2003) Data-gathering wireless sensor networks: organization and capacity. *Comput Netw* 43(4):519–537
2. Gatzianas M, Georgiadis L (Mar. 2008) A distributed algorithm for maximum lifetime routing in sensor networks with mobile sink. *IEEE Trans Wireless Commun* 7(3):984–994
3. Lin X, Shroff NB (2006) Utility maximization for communication networks with multipath routing. *IEEE Trans Autom Control* 51(5):766–781
4. Zhao M, Yang Y (2012) Optimization-based distributed algorithms for mobile data gathering in wireless sensor networks. *IEEE Trans Mob Comput* 11(10):1464–1477
5. Mehrabi A, Kim K (Mar. 2016) Maximizing data collection throughput on a path in energy harvesting sensor networks using a mobile sink. *IEEE Trans Mobile Comput* 15(3):690–704
6. Khodashahi MH, Tashtarian F, Mohammad MHY, Honary MT (2010) Optimal location for mobile sink in wireless sensor networks. In: *Proceedings. IEEE wireless communications and networking conference (WCNC)*, pp 1–6
7. El Monser M, Chikha HB, Attia R (2018) Prolonging the lifetime of large-scale wireless sensor networks using distributed cooperative transmissions. *IET Wireless Sens Syst* 8(5):10
8. Gu X, Zhou X, Yuan B, Sun Y (2018) A Bayesian compressive data gathering scheme in wireless sensor networks with one mobile sink. *IEEE Access* (6):47897–47910
9. Huang C, Zhang R, Cui S (2013) Throughput maximization for the gaussian relay channel with energy harvesting constraints. *IEEE J Sel Areas Commun* 31(8):1469–1479
10. Lee S, Lee S (2010) Optimal transmission methodology for QoS provision of multi-hop cellular network. *J Wirel Netw* 16:1313–1327
11. Abdulla AE, Nishiyama H, Kato N (May 2012) Extending the lifetime of wireless sensor networks: a hybrid routing algorithm. *Comput Commun* 35(9):1056–1063
12. Ma C, He J, Chen HH, Tang Z (2013) Coverage overlapping problems in applications of IEEE 802.15.4 wireless sensor networks. *Conference Proceeding of IEEE Wireless Communication Network*:4364–4369
13. Karenos K, Kalogeraki V (2010) Traffic management in sensor networks with a mobile sink. *IEEE Trans Parallel Distrib Syst* 21(10):1515–1530

14. Safdar V, Bashir F, Hamid Z, Afzal H, Pyun JY (2012) A hybrid routing protocol for wireless sensor networks with mobile sinks. In: Proceedings of 7th International Symposium on Wireless Pervasive Computing, pp 1–5
15. Nagarajan M, Karthikeyan S (2012) A new approach to increase the life time and efficiency of wireless sensor network. International Conference on Pattern Recognition, Informatics and Medical Engineering IEEE explorer, March 2012
16. Fu L, Cheng P, Gu Y, Chen J, He T (2013) Minimizing charging delay in wireless rechargeable sensor networks. IEEE INFOCOM

Chapter 88

Why Feature Selection in Data Mining Is Prominent? A Survey



M. Durairaj and T. S. Poornappriya

Abstract Feature selection is employed to diminish the number of features in various applications where data has more than hundreds of attributes. Essential or relevant attribute recognition has converted a vital job to utilize data mining algorithms efficiently in today's world situations. Current feature selection techniques primarily concentrate on obtaining relevant attributes. This paper presents the notions of feature relevance, redundancy, evaluation criteria, and literature survey on the feature selection approaches in the different areas by many researchers. This paper supports to choose feature selection techniques without identifying the knowledge of every algorithm.

Keywords Relevance · Redundancy · Feature selection · Filter-based approach · Wrapper-based approach · Classification techniques

Abbreviations

AA	Average Accuracy
AIC	Akaike information criterion
ANN	Artificial Neural Network
AUC	Area under the Curve
BWO	Binary Wolf Optimization
CART	Classification and Regression Tree
CFA	Cuttlefish algorithm
CFS	Correlation-based Feature Selection
CS	Chi-Square
DM	Data Mining
F	F-Score

M. Durairaj · T. S. Poornappriya (✉)
School of Computer Science, Engineering and Applications, Bharathidasan University,
Tiruchirappalli, Tamil Nadu, India

FCBF	Fast Correlation-based Feature selection
FP	False Positive
GA	Genetic Algorithm
GR	Gain Ratio
IG	Information Gain
K-NN	K-Nearest Neighbor
LMT	Logistic Model Tree
MI	Mutual Information
MLP	Multi-Layer Perceptron
NB	Naïve Bayes
OA	Overall Accuracy
P	Precision
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
R	Recall
RBF	Radial Basis Function
ROC	Receiver Operating Curve
SVM	Support Vector Machine
TP	True Positive
TV	Term Variance
WOA	Whale Optimization algorithm

88.1 Introduction

The amount of high-dimensional data that persists and is available on the Internet has dramatically improved in the prior few years. Hence, machine learning techniques have the difficulty of handling with the critical number of input attributes, which is creating an essential problem for researchers. Data preprocessing has been necessitated for utilizing Machine Learning methods efficiently. Feature selection [17] is an example of the prominent and most common approaches in data preprocessing and has converted a significant element of the machine learning method; it is also called an attribute selection, variable subset selection, or variable selection in statistics and machine learning. It is the process of eliminating unnecessary and identifying redundant, relevant features or noisy data. This method expedites data mining algorithms and improves predictive accuracy and understandability. Irrelevant attributes are those that provide no valuable information, and unnecessary features contribute no extra information than the presently selected attributes. About supervised inductive learning, attribute selection shows a collection of candidate attributes utilizing one of the three procedures.

- An evaluation metric is optimized by the specific size of the attribute subset.
- A particular constraint on evaluation metric is satisfied by the diminutive size of the attribute subset.

- The best commitment is the subset with evaluation metric and size.

In the method of feature selection [1], redundant or noise attributes in the data may be back in various situations, because they are not appropriate and necessary for the class notion such as microarray data investigation. When the amount of individuals is hugely smaller than the attribute, then machine learning becomes especially tricky because the search space will be lightly populated. Therefore, the design will not be capable of distinguishing precisely between relevant and noise data. There are two principal approaches to feature selection. Among these, subset evaluation is the primary, and the second one suggests Individual Evaluation. In the Individual Evaluation, the weight of an individual attribute is allocated according to its degree of relevance. In Subset Evaluation, candidate attribute subsets are formed utilizing a search approach.

88.2 Importance of Feature Selection

Several feature selection techniques have been proposed in the paper, and their comparative examination is a challenging task. Without identifying the appropriate attributes in an approach of the live dataset, it is complicated to determine the strength of the feature selection techniques because datasets may incorporate several hurdles such as the vast number of redundant and irrelevant, high dimensionality and noisy data in terms of samples or features. Hence, the efficiency of the feature selection technique swears on the learning approach performance. There are several performance criteria discussed in the literature such as the ratio of feature selection, accuracy, error rates, etc. Most researchers accept that there is no so-called “trust-worthy system.” Therefore, the new feature selection techniques are continually developing to undertake the particular difficulty with various procedures. It includes utilizing an ensemble method for assuring a reliable performance of feature selection [37].

88.3 Defining Relevancy and Redundancy of Features

88.3.1 Feature Relevancy

The attributes in the dataset have been classified into three dissociate categories called irrelevant attributes, strongly relevant, and weakly relevant. Let A be a full set of attributes, A_j a feature, and $SS_j = A - \{A_j\}$. These classes of relevance are validated as follows:

Definition 1: Strong Relevance A feature A_j is strongly relevant if and only if

$$P(C | A_j, SS_j) \neq P(C | SS_j)$$

The strong relevance of an attribute shows that the feature is always essential for an optimal subset; it cannot be separated without changing the above conditional class configuration.

Definition 2: Weak Relevance A feature A_j is weakly relevant if and only if

$$P(C | A_j, SS_j) = P(C | A_j), \text{ and}$$

$$\exists SS'_j \subset SS_j, \text{ such that } P(C | A_j, SS'_j) \neq P(C | SS'_j).$$

Weak relevance recommends that the attributes may not always be required but may become essential for an optimal subset at specific circumstances.

Definition 3: Irrelevance An attribute A_j is irrelevant if and only if

$$\forall SS'_j \subseteq SS_j, P(C | A_j, SS'_j) = P(C | SS'_j)$$

Irrelevance (following Definitions 1 and 2) means that the attribute is not significant at all.

88.3.2 Feature Redundancy

Assumptions of attribute redundancy usually are regarding attribute correlation. It has broadly affirmed that two attributes are unnecessary to each other if their conditions are entirely correlated. In reality, it may not be so sincere to discover attribute redundancy when an attribute is correlated (conceivably partially) with a set of attributes.

Definition 4: Markov Blanket Given a feature A_j , let $B_j \subset A$ ($A_j \notin B_j$), M is said to be a Markov blanket for A_j if and only if

$$P(A - B_j - \{A_j\}, C | A_j, B_j) = P(A - B_j - \{A_j\}, C | B)$$

The Markov blanket circumstance needs that B_j associate not only the information that A_j has about C , but also about all of the other attributes.

Definition 5: Redundant Features Let R be the current collection of attributes, an attribute is redundant and hence should be eliminated from R if and only if it is weakly relevant and has a Markov blanket B_j within R .

88.4 Literature Survey on Feature Selection Techniques in Various Applications

88.4.1 Literature Survey on Feature Selection in the Year 2015

The authors Sung-Sam Hong in [1] used Genetic Algorithm (GA) for selecting a set of optimized features from the corpus. GA is used for the text classification problems.

The authors Yuhua Qian in [2] used fuzzy rough feature selection algorithm for time consumption and to analyze the datasets with large scale and high dimensionality. The authors used Fuzzy positive region reduction and Fuzzy information entropy reduction for the high-dimensional datasets from University of California, Irvine (UCI) repository.

The authors Deron Liang in [3] used many techniques like Linear Discriminant Analysis, t-Test, Logistic regression, Genetic algorithms, and Particle Swarm Optimization for predicting the financial distress.

The authors Emina in [4] used Genetic algorithm as a feature selector for predicting the breast cancer and it is evaluated by various classification algorithms with different performance metrics.

The author Kusum in [5] used Term variance (TV), Document Frequency, and Principal Component Analysis (PCA) in the preprocessing steps for feature selection. The authors applied these algorithms in the text classification problem.

The author Hannah in [6] used Rough Set theory and Evolutionary methods for predicting the diseases using difference medical dataset. This paper is composed of Rough Set theory, Quick Reduct, Particle Swarm Optimization, and Harmony search algorithm for feature selection.

The author Chan in [7] used Random forest classifier as the feature selector for medical datasets and its efficiency is evaluated by K-Nearest Neighbor classifier with different evaluation metrics.

The authors Min Han in [8] utilized filter-based approach and wrapper-based approach feature selection for the synthetic dataset of Friedman and the real life dataset of meteorological time series. In this paper, the researchers applied Binary Particle Swarm Optimization (PSO) and Mutual information, where, for feature selection, Genetic Algorithm is utilized.

The authors Fatemeh in [9] used principal component analysis (PCA), genetic algorithm (GA), information gain ratio, and relief feature selection methods for credit scoring. These feature selection methods are evaluated with classification techniques and different evaluation metrics.

The authors Adel in [10] suggested Cuttlefish algorithm (CFA) for intrusion detection system. The performance of the CFA has been evaluated by Decision Tree classification with Detection rate, Accuracy, and False Alarm rate metrics. The authors Nicoletta in [11] utilized filter-based feature selection for three different

problems. The techniques like Chi-Squared, Information Gain, Symmetrical Uncertainty, Gain Ratio, OneR, and ReliefF feature selection are used in this paper. The efficiency of these techniques is evaluated by Support Vector Machine classification method.

88.4.2 Literature Survey on Feature Selection in the Year 2016

Asha in [12], the authors utilized Gini Index as the feature selector for sentiment analysis of the movie review problems. This paper used SVM as the classifier.

Opeyemi in [13], the authors used filter-based feature selection methods like IG, gain ratio, chi-squared, and ReliefF for detecting the Distributed Denial of Service in cloud environment. Decision Tree classifications like C5.0 and C4.5 methods are used for evaluating the feature selection methods.

Farideh in [14], the authors used both filter- and wrapper-based feature selection methods for the clinical prediction models and diabetes prediction. These methods are analyzed by Random forest classification method with different performance metrics like Akaike information criterion (AIC) and area under the curve (AUC).

E. Emary in [15], the authors introduced a new optimization technique called Grey Wolf optimization and Binary BGWO as the feature selector in the preprocessing stage. These two techniques are evaluated by K-Nearest Neighbor (KNN) classification method.

Abdullah in [16], the authors used Genetic Algorithm for the problem of text categorization and text classification in text mining.

Youchuan in [17], the authors used optimization techniques like Genetic algorithm, Ant Colony Optimization, and Particle Swarm Optimization as the feature selection methods in the preprocessing stage. These optimization techniques are evaluated by KNN and SVM with 10 different UCI repository datasets.

Maolong Xi in [18], the authors utilized Particle Swarm Optimization and Binary Quantum-behaved PSO as the feature selector for the cancer classification problems.

Liming Shen in [19], authors suggested a new optimization algorithm called Fruit Fly Optimization algorithm for the medical data classification problems. This proposed optimization method was evaluated by three different classification techniques called Support Vector Machine (SVM), GA-SVM, and PSO-SVM.

Kusum in [20], the authors utilized Binary Particle Swarm Optimization for text classification problems. For evaluating the feature selection techniques, two classification methods like Naïve Bayes and Support Vector Machine have been utilized.

Javier in [21], authors used Binary Differential Evolution, Rank-based filter feature selection method for high-dimensional microarray experiments. Naïve Bayes, Support Vector Machine, C4.5, and K-Nearest Neighbor classification techniques are used for evaluating the feature selection with different performance metrics.

88.4.3 Literature Survey on Feature Selection in the Year 2017

The authors Majdi in [22] introduced a new feature selection method by combining Whale Optimization algorithm and Simulated Annealing. It is tested with 18 Standard Benchmark Datasets considered and K-NN classification method.

The authors Hossam Faris in [23] used Multi-verse optimizer as a feature selector and it is evaluated in 10 benchmark datasets and Support Vector Machine classifier.

The authors Sumaiya in [24] utilized Chi-Square, Radial Basis Function (RBF), and Kernel function for the feature selection in the intrusion detection. It is evaluated by Support Vector Machine classifier with various performance metrics.

Laith [25] applied Genetic Algorithm and Particle Swarm Optimization as the feature selection technique for text clustering and classification.

Shunmugapriya [26] utilized Ant Colony Optimization and Bee Colony Optimization for feature selection. These two techniques have been executed as the feature selector for evaluation by 13 UCI repository datasets.

Alaa [27], authors suggested Bat algorithm, genetic algorithm, and particle swarm optimization as the feature selection method. These techniques have been estimated with SVM classifier by nine UCI repository datasets.

Chengming [28], authors presented standard deviation, Kullback–Leibler distance, Correlation Coefficient, and Particle Swarm Optimization for the Hyperspectral classification. SVM was used as the classifier for evaluation.

Prashant [29] utilized optimization techniques like Particle Swarm Optimization, Binary Bat Algorithm, Genetic algorithm, and Modified Cuckoo search algorithm as the feature selection methods in the preprocessing stage for the prediction of Parkinson's disease.

Worawut [30], authors presented Particle Swarm Optimization for the diagnosis of leukemia with Support Vector Machine classifier.

Hong Wang [31], authors introduced Bacterial Colony Optimization utilized for feature selection in the Microarray Gene Expression Cancer Data with the K-Nearest Neighbor algorithm (KNN), SVM classifiers.

88.4.4 Literature Survey on Feature Selection in the Year 2018

Shenkai Gu in [32], the authors introduced Competitive Swarm Optimizer and Particle Swarm Optimization as the feature selection methods. Nine benchmark UCI repository datasets have been utilized for evaluating these algorithms. For assessing the feature selection techniques, the KNN classification method has been utilized.

Shadi in [33], the authors used Filter-based feature selection methods for Intrusion Detection System. Information Gain and Mutual Information methods are utilized as the feature selector.

The authors Emrah in [34] presented Artificial bee colony, nondominated sorting procedure, and genetic operators as the feature selector. These methods have been assessed and examined with 12 UCI repository datasets.

Majdi in [35], the authors introduced Grasshopper Optimization Algorithm as the feature selector. This method was evaluated with K-Nearest Neighbor classification technique and 22 UCI repository datasets.

Neha in [36], the authors suggested an intelligent water drop algorithm to be used for intrusion detection system. Support Vector Machine was used as the classifier for evaluation of the proposed feature selection method.

The authors Ramalingaswamy in [37] used Rough Set theory, Bat optimization, Quick Reduct, and Fuzzy Rules in this paper for the feature selection process. With these feature selection methods, AdaBoosting classification method is used for detection of diabetes disease with various performance metrics.

Indu in [38], the authors introduced filter-based and wrapper-based methods for the classification of cancer and gene selection. Particle Swarm Optimization (PSO) and Correlation-based feature selection approaches have been utilized as the feature selector with Naïve Bayes classifier.

The authors Mao-Te in [39] suggested Gain Ratio filter-based feature selection method for predicting the length of stay in hospital for patients. Different classification techniques like Logistic Regression, CART, Random Forest, C4.5, and Support Vector Machine are used for the evaluation of the feature selection technique.

Following Table 88.1 summarizes the literature survey on the feature selection methodologies by using Data Mining techniques. This Table 88.1 gives the literature survey on the year 2015–2018.

88.5 Outcome of the Literature Survey

From the above literature survey on the feature selection using Data Mining techniques, it is clear that the feature selection can be done by any one of the following methods with performance metrics.

Filter-Based Feature Selection Methods The filter approach actually precedes the actual classification process. The filter-based feature selection is computationally scalable, simple, and fast. Correlation-based Feature Selection (CFS), Principal Component Analysis (PCA), Gain Ratio (GR), Information Gain (IG), Chi-Squared analysis (CS), ReliefF, OneR, Fast Correlation-based Feature selection (FCBF), Euclidean distance, i-test, Markov blanket filter. Often, the filters are used with forward selection, backward elimination, bidirectional search, best-first search, genetic search, and other methods.

Table 88.1 Literature survey on feature selection using data mining techniques

Reference	Feature selection method used	Classifier/ clustering used	Domain	Performance metrics considered
[1]	Genetic algorithm	Hierarchical clustering and K-means clustering	Text classification problem	Accuracy, F1-measure
[2]	Fuzzy positive region reduction, fuzzy information entropy reduction	–	Six benchmark datasets	Time consumption
[3]	Linear discriminant analysis, t-test, logistic regression, genetic algorithms, particle swarm optimization	Linear SVM, RBF, SVM, K-NN, Naïve Bayes, CART, and MLP	Financial distress prediction	Type 1 error, accuracy
[4]	Genetic algorithm	Logistic regression, Bayesian network, MLP, Radial Basis Function Network (RBFN), SVM, C4.5 decision tree, random Forest	Breast cancer detection	Accuracy, area under the curve (AUC), F-measure
[5]	Term variance (TV), document frequency, principal component analysis (PCA)	k-means clustering algorithm	Text clustering	Precision, recall, F1-measure
[6]	Rough set theory, quick Reduct, particle swarm optimization, harmony search algorithm	Logistic model tree (LMT)	Medical datasets	Precision, recall, F-measure, overall accuracy, mean absolute error
[7]	Feature selection based on random Forest classifier	K-nearest neighbor	Medical datasets	Accuracy, number of selected features, standard deviation
[8]	Mutual information, binary PSO, genetic algorithm	–	The synthetic dataset of Friedman and the real life dataset of meteorological time series	Root mean square error, symmetric mean absolute percentage error
[9]	Principal component analysis (PCA), genetic algorithm (GA), information gain ratio, and relief	SVM, ANN, AdaBoost, CART, NB, random Forest	Credit scoring	Accuracy, area under the curve (AUC), number of selected features
[10]	Cuttlefish algorithm (CFA)	Decision tree classification	Intrusion detection system	Detection rate, accuracy, false alarm rate

(continued)

Table 88.1 (continued)

Reference	Feature selection method used	Classifier/ clustering used	Domain	Performance metrics considered
[11]	Chi-squared, information gain, symmetrical uncertainty, gain ratio, OneR, ReliefF	Support vector machine	Internet advertisements, text categorization, microarray data classification	Kuncheva similarity
[12]	Gini index	Support vector machine	Sentiment analysis for movie review dataset	Error rates and accuracy
[13]	Information gain (IG), gain ratio, chi-squared, And ReliefF	Decision tree classifications like C5.0 and C4.5	Distributed denial of service detection in cloud computing	Detection rate and classification accuracy
[14]	Filter and wrapper methods	Random forest method	Clinical prediction models: Diabetes prediction	Akaike information criterion (AIC) and area under the curve (AUC)
[15]	Grey wolf optimization (GWO), binary GWO	K-nearest neighbor (KNN)	18 UCI repository datasets	Classification accuracy, statistical best, statistical mean, statistical worst, standard mean, average selection size, average F-score.
[16]	Genetic algorithm	Naïve Bayes, associative classification	Text categorization and classification	Reduction rate
[17]	Genetic algorithm, ant colony optimization, particle swarm optimization	K-nearest neighbor algorithm (KNN), SVM	10 datasets from UCI repository	Accuracy, fitness value, central processing unit (CPU) time, number of selected features
[18]	Particle swarm optimization (PSO), binary quantum-behaved PSO	SVM	Cancer classification	Accuracy, correct rate, search capability
[19]	Fruit fly optimization algorithm	SVM, GA-SVM, PSO-SVM	Medical data classification	Classification accuracy, sensitivity, specificity, AUC (the area under the receiver operating characteristic (ROC) curve), and processing time.

(continued)

Table 88.1 (continued)

Reference	Feature selection method used	Classifier/ clustering used	Domain	Performance metrics considered
[20]	Binary particle swarm optimization	Support vector machine, Naïve Bayes	Text classification	Precision (P), recall (R), and F-score (F)
[21]	Binary differential evolution, rank-based filter feature selection (FS) method	NB, SVM, C4.5, and KNN	High-dimensional microarray experiments	True positive, false positive, ROC
[22]	Whale optimization algorithm (WOA), simulated annealing	K-nearest neighbor classification	18 standard benchmark datasets considered	Classification accuracy, number of features selected
[23]	Multi-verse optimizer	SVM	Ten benchmark datasets	Accuracy, number of features selected
[24]	Chi-Square, RBF, kernel function	Support vector machine	Intrusion detection system	Classification accuracy, kappa statistic, mean absolute error, root mean squared error, true positive (TP) rate, false positive (FP) rate, precision, recall, F-measure, ROC area
[25]	Particle swarm optimization, genetic algorithm	k-means clustering	Text classification and clustering	Computation time, accuracy, precision, recall, and F-measure
[26]	Ant Colony optimization, bee Colony optimization	SVM, neural network	13 UCI repository datasets	Classification accuracy, computation complexity, F-test
[27]	Bat algorithm, genetic algorithm, particle swarm optimization	Support vector machine	9 UCI repository datasets	Accuracy, classification error rates
[28]	Standard deviation, Kullback–Leibler distance, correlation coefficient, particle swarm optimization	Support vector machine	Hyperspectral classification	Overall accuracy (OA), average accuracies (AA), and kappa coefficient
[29]	Particle swarm optimization, binary bat	Artificial neural network	Parkinson’s disease prediction	Accuracies, true positive rates, false positive

(continued)

Table 88.1 (continued)

Reference	Feature selection method used	Classifier/ clustering used	Domain	Performance metrics considered
	algorithm, genetic algorithm, modified cuckoo search algorithm			rates, positive predicted values, And negative predicted values
[30]	Particle swarm optimization	Support vector machine	Leukemia diagnosis	Accuracy, number of selected features
[31]	Bacterial colony optimization	K-nearest Neighbor algorithm (KNN), SVM	Microarray gene expression Cancer data	Accuracy, number of selected features
[32]	Particle swarm optimization, competitive swarm optimizer	K-nearest neighbor	Nine benchmark UCI repository datasets	Error rate, number of selected features
[33]	Information gain (IG), mutual information (MI)	J48, random tree, Naïve Bayes	Intrusion detection system	True positive rate, accuracy, false positive rate, ROC area
[34]	Artificial bee colony, nondominated sorting procedure, And genetic operators	–	12 UCI repository datasets	Accuracy, number of selected features
[35]	Grasshopper optimization algorithm	K-nearest neighbor	22 UCI repository datasets	Accuracy, standard deviation, number of selected attributes, CPU time
[36]	Intelligent water drop algorithm	Support vector machine	Intrusion detection system	Detection rate, false alarm rate, accuracy
[37]	Rough set theory, bat optimization, quick Reduct, fuzzy rules	AdaBoosting	Diabetes disease detection	Computation time, number of selected features, accuracy, sensitivity, specificity, G-measure, true positive, true negative, false negative, false positive
[38]	Correlation-based feature selection, particle swarm optimization	Naïve Bayes	Gene selection and Cancer classification	Accuracy, standard deviation, mean, execution time
[39]	Gain ratio	Logistic regression, CART, random forest, C4.5, SVM	Length of stay in hospital prediction	Accuracy, specificity, sensitivity, and AUC

Wrapper-Based Feature Selection Wrapper model approach uses the method of classification itself to measure the importance of features set; hence, the feature selected depends on the classifier model used. Evolutionary algorithms like Particle Swarm Optimization, Ant Colony Optimization, Genetic Algorithm, Bee Colony Optimization, etc. Among these evolutionary algorithms, PSO is widely used in the process of feature selection. However, wrapper methods are too expensive for large-scale dimensional datasets in terms of time and computational complexity.

Performance Metrics The widely used parameters to evaluate the efficiency of the feature selection techniques by many researchers are Accuracy, Number of Features Selected, Error rates, Precision, Recall, True Positive, False Positive, and ROC curve.

88.6 Research Directions

Through this literature survey, it has been concluded that the proposed feature selection must decide the low redundant and utterly relevant attributes from the dataset. Several machine learning algorithms have been applied for estimating the performance of the proposed feature selection. Following are the future research directions obtained by this detailed survey of feature selection techniques in various domains.

- Generating a unifying structure for information-theoretic feature selection.
- Enhancing the efficiency and efficacy of information-theoretic feature selection techniques in high-dimensional ranges.
- Examining the relationship between the filter-based feature selection methods and Machine Learning classification techniques for better prediction.
- Developing a framework for analyzing the relationship between causal discovery and feature selection.
- Expanding new measures of statistical confidence beyond Information-Theoretic feature selection methods and correlation.

88.7 Conclusion

From this survey paper, it is clear that the data dimension reduction or variable screening or feature selection in predictive analytics belongs to the method of recognizing the few common essential parameters or variables which supports in prognosticating the result. Feature selection is the method of narrowing down a subset of attributes, or features, to be utilized in the predictive prototyping method. Feature selection is valuable on a mixture of aspects: it is the best defense upon the Curse of Dimensionality; it can diminish overall training events; and it is a defense mechanism toward overfitting, developing model generalizability.

References

1. Hong S-S, Lee W, Han M-M (2015) The feature selection method based on genetic algorithm for efficient of text clustering and text classification. *Int J Adv Soft Comput Appl* 7:1
2. Qian Y et al (2015) Fuzzy-rough feature selection accelerator. *Fuzzy Sets Syst* 258:61–78
3. Liang D, Tsai C-F, Wu H-T (2015) The effect of feature selection on financial distress prediction. *Knowl-Based Syst* 73:289–297
4. Aličković E, Subasi A (2017) Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Comput & Applic* 28(4):753–763
5. Bharti KK, Singh PK (2015) Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering. *Expert Syst Appl* 42(6):3105–3114
6. Inbarani HH, Bagyamathi M, Azar AT (2015) A novel hybrid feature selection method based on rough set and improved harmony search. *Neural Comput & Applic* 26(8):1859–1880
7. Park CH, Kim SB (2015) Sequential random k-nearest neighbor feature selection for high-dimensional data. *Expert Syst Appl* 42(5):2336–2342
8. Han M, Ren W (2015) Global mutual information-based feature selection approach using single-objective and multi-objective optimization. *Neurocomputing* 168:47–54
9. Koutanaei FN, Sajedi H, Khanbabaie M (2015) A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *J Retail Consum Serv* 27:11–23
10. Eesa AS, Orman Z, Brifciani AMA (2015) A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems. *Expert Syst Appl* 42(5):2670–2679
11. Dessì N, Pes B (2015) Similarity of feature selection methods: An empirical study across data intensive classification tasks. *Expert Syst Appl* 42(10):4632–4642
12. Manek AS et al (2017) Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web* 20(2):135–154
13. Osanaiye O et al (2016) Ensemble-based multi-filter feature selection method for DDoS detection in cloud computing. *EURASIP J Wirel Commun Netw* 1(2016):130
14. Bagherzadeh-Khiabani F et al (2016) A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. *J Clin Epidemiol* 71:76–85
15. Emary E, Zawbaa HM, Hassanien AE (2016) Binary grey wolf optimization approaches for feature selection. *Neurocomputing* 172:371–381
16. Ghareb AS, Bakar AA, Hamdan AR (2016) Hybrid feature selection based on enhanced genetic algorithm for text categorization. *Expert Syst Appl* 49:31–47
17. Wan Y et al (2016) A feature selection method based on modified binary coded ant colony optimization algorithm. *Appl Soft Comput* 49:248–258
18. Xi M et al (2016) Cancer feature selection and classification using a binary quantum-behaved particle swarm optimization and support vector machine. *Comput Math Methods Med* 2016
19. Shen L et al (2016) Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl-Based Syst* 96:61–75
20. Bharti KK, Singh PK (2016) Opposition chaotic fitness mutation based adaptive inertia weight BPSO for feature selection in text clustering. *Appl Soft Comput* 43:20–34
21. Apolloni J, Leguizamón G, Alba E (2016) Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments. *Appl Soft Comput* 38:922–932
22. Mafarja MM, Mirjalili S (2017) Hybrid Whale Optimization Algorithm with simulated annealing for feature selection. *Neurocomputing* 260:302–312
23. Faris H et al (2017) A multi-verse optimizer approach for feature selection and optimizing SVM parameters based on a robust system architecture. *Neural Comput & Applic*:1–15
24. Thaseen IS, Kumar CA (2017) Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *J King Saud Univer Comp Inform Sci* 29(4):462–472

25. Abualigah LM, Khader AT (2017) Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering. *J Supercomput* 73(11):4773–4795
26. Shunmugapriya P, Kanmani S (2017) A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid). *Swarm Evol Comput* 36:27–36
27. Tharwat A, Hassanien AE, Elnaghi BE (2017) A ba-based algorithm for parameter optimization of support vector machine. *Pattern Recogn Lett* 93:13–22
28. Qi C et al (2017) Feature selection and multiple kernel boosting framework based on PSO with mutation mechanism for hyperspectral classification. *Neurocomputing* 220:181–190
29. Shrivastava P et al (2017) A survey of nature-inspired algorithms for feature selection to identify Parkinson's disease. *Comput Methods Program Biomed* 139:171–179
30. Srisukkhom W et al (2017) Intelligent leukaemia diagnosis with bare-bones PSO based feature optimization. *Appl Soft Comput* 56:405–419
31. Wang H, Niu B (2017) A novel bacterial algorithm with randomness control for feature selection in classification. *Neurocomputing* 228:176–186
32. Gu S, Cheng R, Jin Y (2018) Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Comput* 22(3):811–822
33. Aljawarneh S, Aldwairi M, Yassein MB (2018) Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model. *J Comput Sci* 25:152–160
34. Hancer E et al (2018) Pareto front feature selection based on artificial bee colony optimization. *Inf Sci* 422:462–479
35. Mafarja M et al (2018) Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems. *Knowl-Based Syst* 145:25–45
36. Acharya N, Singh S (2018) An IWD-based feature selection method for intrusion detection system. *Soft Comput* 22(13):4407–4416
37. Cheruku R et al (2018) RST-BatMiner: a fuzzy rule miner integrating rough set feature selection and Bat optimization for detection of diabetes disease. *Appl Soft Comput* 67:764–780
38. Jain I, Jain VK, Jain R (2018) Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. *Appl Soft Comput* 62:203–215
39. Chuang M-T, Hu Y-h, Lo C-L (2018) Predicting the prolonged length of stay of general surgery patients: a supervised learning approach. *Int Trans Oper Res* 25(1):75–90