Fatima Dargam
Pavlos Delias
Isabelle Linden
Bertrand Mareschal (Eds.)

LNBIP 313

Decision Support Systems VIII

Sustainable Data-Driven and Evidence-Based Decision Support

4th International Conference, ICDSST 2018
Heraklion, Greece, May 22–25, 2018
Proceedings

EWG DSS 2018 ICDSST PROMETHEE 2018 days

Springer

Lecture Notes in Business Information Processing 313

More information about this series at http://www.springer.com/series/7911

Fatima Dargam · Pavlos Delias
Isabelle Linden · Bertrand Mareschal (Eds.)

Decision Support Systems VIII

Sustainable Data-Driven and Evidence-Based Decision Support

4th International Conference, ICDSST 2018
Heraklion, Greece, May 22–25, 2018
Proceedings

Springer

Editors
Fatima Dargam
SimTech Simulation Technology
Graz
Austria

Pavlos Delias
Eastern Macedonia and Thrace Institute of Technology
Kavala
Greece

Isabelle Linden
University of Namur
Namur
Belgium

Bertrand Mareschal
ULB
Brussels
Belgium

ISSN 1865-1348 ISSN 1865-1356 (electronic)
Lecture Notes in Business Information Processing
ISBN 978-3-319-90314-9 ISBN 978-3-319-90315-6 (eBook)
https://doi.org/10.1007/978-3-319-90315-6

Library of Congress Control Number: 2018940641

Printed on acid-free paper

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

EURO Working Group on Decision Support Systems

The EWG-DSS is a Euro Working Group on Decision Support Systems within EURO, the Association of the European Operational Research Societies. The main purpose of the EWG-DSS is to establish a platform for encouraging state-of-the-art high-quality research and collaboration work within the DSS community. Other aims of the EWG-DSS are to:

- Encourage the exchange of information among practitioners, end-users, and researchers in the area of decision systems
- Enforce the networking among the DSS communities available and facilitate activities that are essential for the start up of international cooperation research and projects
- Facilitate the creation of professional, academic, and industrial opportunities for its members
- Favor the development of innovative models, methods, and tools in the field of decision support and related areas
- Actively promote the interest in decision systems in the scientific community by organizing dedicated workshops, seminars, mini-conferences, and conferences, as well as editing special and contributed issues in relevant scientific journals

The EWG-DSS was founded with 24 members, during the EURO Summer Institute on DSS that took place at Madeira, Portugal, in May 1989, organized by two well-known academics of the OR community: Jean-Pierre Brans and José Paixão. The EWG-DSS group has substantially grown along the years. Currently, we have over 300 registered members from around the world.

Through the years, much collaboration among the group members has generated valuable contributions to the DSS field, which resulted in many journal publications. Since its creation, the EWG-DSS has held annual meetings in various European countries, and has taken active part in the EURO Conferences on decision-making-related subjects. Starting from 2015, the EWG-DSS established its own annual conferences, namely, the International Conference on Decision Support System Technology (ICDSST).

The current EWG-DSS Coordination Board comprises seven experienced scholars and practitioners in the DSS field: Pascale Zaraté (France), Fátima Dargam (Austria), Shaofeng Liu (UK), Boris Delibašic (Serbia), Isabelle Linden (Belgium), Jason Papathanasiou (Greece), and Pavlos Delias (Greece).

Preface

The proceedings of the eighth edition of the EWG-DSS Decision Support Systems published in the LNBIP series present a selection of reviewed and revised full papers from the 4th International Conference on Decision Support System Technology and PROMETHEE days (ICDSST – PROMETHEE DAYS 2018) held in Heraklion, Greece, during May 22–25, 2018, with the main theme: "Sustainable Data-Driven and Evidence-Based Decision Support with Applications to the Environment and Energy Sector." This event was jointly organized by the EURO Working Group on Decision Support Systems (EWG-DSS) and the PROMETHEE DAYS group and it was hosted by the Hellenic Centre for Marine Research.

The EWG-DSS series of the International Conference on Decision Support System Technology (ICDSST), starting with ICDSST 2015 in Belgrade, was planned to consolidate the tradition of annual events organized by the EWG-DSS in offering a platform for European and international DSS communities, comprising the academic and industrial sectors, to present state-of-the-art DSS research and developments, to discuss current challenges that surround decision-making processes, to exchange ideas about realistic and innovative solutions, and to co-develop potential business opportunities.

The scientific topic areas of ICDSST 2018 include:

- Decision support modeling
- Decision support and expert systems in the environment and energy sectors
- Data visualization and decision-making/support
- Environmental and energy data visualization
- Geographic information systems
- Data science, big data analytics, and business intelligence
- Technical and managerial dashboards for decision support
- Knowledge acquisition, management, extraction, and visualization
- Collaborative decision-making
- Intelligent, Web-based, spatial decision support systems
- Application models and systems: agricultural, environmental, farming, water resources, livestock, fisheries, warehouse, sustainable forest and rangeland management, food processing, and supply chain management
- Simulation models and systems, regional planning, logistics, and traceability
- Case studies involving any of the above issues

The aforementioned topics reflect some of the essential topics in the field of decision support systems, and they represent several topics of the research interests of the group members. This rich variety of themes, advertised not only to the (more than 300) members of the group, but to a broader audience as well, allowed us to gather several contributions regarding the implementation of decision support processes, methods, and technologies in a large variety of domains. Hence, this EWG-DSS LNBIP volume published by Springer has considered contributions of a “full-paper” format, selected through a single-blind paper reviewing process. In particular, at least two reviewers – members of the Program Committee – reviewed each submission through a rigorous two-stage process. Finally, we selected 15 out of 71 submissions, corresponding to a 21.1% acceptance rate, to be included in this 8th EWG-DSS LNBIP volume.

We proudly present the selected contributions, organized in four sections:

1. Decision Support Systems for a Sustainable Society: Suggestions and empirical proofs about how DSS can improve the world we live in and exploit evidence to advance the quality of the decisions made. First, Tong Wang, Qi Han, and Bauke De Vries present “SIRPSS-Sustainable Industrial Site Redevelopment Planning Support System.” Then, “An Intelligent Multi-Agent System Using Fuzzy Analytic Hierarchy Process and Axiomatic Design as a Decision Support Method for Refugee Settlement Siting” by Maria Drakaki, Hacer Guner Goren, and Panagiotis Tzionas deals with a crucial topic of our era, followed by the work of Khelifa Boudjemaa, Ridda Laouar, and Sean Eom about an “Intelligent Integrated System for Urban Planning Using GIS and Cloud Computing.” This section closes with “An Ontology-Based Decision Support Framework for Personalized Quality of Life Recommendations” by Marina Riga, Efstratios Kontopoulos, Kostas Karatzas, Stefanos Vrochidis, and Ioannis Kompatsiaris.
2. Decision Support Systems Serving the Public: Methods and case studies on DSS in the service of the public, with topics ranging from physical disaster protection to policy design. Stefano Armenia, Georgios Tsaples, and Camillo Carlini present “Critical Events and Critical Infrastructures: A System Dynamics Approach” and Marta Dell'Ovo, Francesca Torrieri, and Alessandra Oppio present “How to Model Stakeholder Participation for Flood Management.” This section continues with a case study from Brazil on how “Big Data Analytics Improve the Decision-Making Process in Public Safety,” a work by Jean Turet and Ana Paula Cabral Seixas Costa, and finishes with the work of Glykeria Myrovali, Georgios Tsaples, Maria Morfoulaki, Georgia Aifadopoulou, and Jason Papathanasiou about “An Interactive Learning Environment Based on System Dynamics Methodology for Sustainable Mobility Challenges Communication and Citizens’ Engagement.”
3. Decision Support Systems in Management and Organizations: A section that highlights how DSS can be involved in management practices of modern organizations to give a boost to their effectiveness and efficiency. Yusuf-Ali Karbelkar and Mike Hart present “Skills and Mindsets for an Analytically Innovative Organization,” while Georgios Stavrou, Panagiotis Adamidis, and Jason Papathanasiou deal with “Computer Supported Team Formation.” Then, the important as well as frequent problem of supplier selection is presented through a description of the case of a “Colombian Agricultural Research Company,” a work by Jenny Milena Moreno Rodriguez, Takanni Hannaka Abreu Kang, Eduarda Asfora Frej, and Adiel Teixeira de Almeida.
4. Advances in Decision Support Systems Technologies and Methods: Abstracting from the thematic focus of the conference, this section focuses on methods, techniques, approaches, and technologies that advance the research of the DSS field. Dario Torregrossa and Joachim Hansen present “SK-DSSy: How to Integrate the

YouTube Platform in a Cooperative Decision Support," while a valuable family of methods in DSS, multiple-criteria decision aid, is exemplified through the work of Gjorgji Nusev, Pavle Boškoski, Marko Bohanec, and Biljana Mileva Boshkoska: "A DSS Model for Selection of Computer-on-Module Based on PROMETHEE and DEX Methods." Then, Kathrin Kirchner and Petar Markovic present how process mining can be related to the field of decision support with their work on "Unveiling Hidden Patterns in Flexible Medical Treatment Processes – A Process Mining Case Study." Finally, "A New Function for Ensemble Pruning" is suggested by Souad Taleb Zouggar and Abdelkader Adla.

We would like to thank many people who contributed greatly to the success of this LNBIP book. First of all, we would like to thank Springer for providing us with the opportunity to guest edit the DSS book, and we wish to express our sincere gratitude to Ralf Gerstner and Christine Reiss, who dedicated their time to guide and advise us during the volume editing process. Secondly, we need to thank all the authors for submitting their state-of-the-art work for consideration for this volume. From our point of view, this is yet another confirmation that the DSS community is vivid, active, and has a great potential for contributions. It really gives us courage and stimulates us to continue the International Conference on DSS Technology series. Finally, we express our deep gratitude to the reviewers and members of the Program Committee who volunteered to assist in the improvement and the selection of the papers, under a tight schedule. We believe that this EWG-DSS Springer LNBIP volume presents a rigorous selection of high-quality papers addressing the conference theme. We hope that readers will enjoy the publication!

March 2018

Fatima Dargam
Pavlos Delias
Isabelle Linden
Bertrand Mareschal

Organization

Conference Chairs

Fátima Dargam	SimTech Simulation Technology, Austria
Boris Delibašić	University of Belgrade, Serbia
Pavlos Delias	Eastern Macedonia and Thrace Institute of Technology, Greece
Isabelle Linden	University of Namur, Belgium
Shaofeng Liu	University of Plymouth, UK
Bertrand Mareschal	Université libre de Bruxelles, Belgium
Jason Papathanasiou	University of Macedonia, Greece
Pascale Zaraté	IRIT/Toulouse 1 Capitole University, France

Program Committee

Irène Abi-Zeid	FSA – Laval University, Canada
Abdelkader Adla	University of Oran 1, Algeria
Carlos Henggeler Antunes	University of Coimbra, Portugal
Francisco Antunes	Beira Interior University, Portugal
George Aretoulis	Aristotle University of Thessaloniki, Greece
Christos Arvanitidis	IMBBC-HCMR, Greece
Georgia Ayfantopoulou	HIT, Greece
Dragana Bečejski-Vujaklija	Serbian Society for Informatics, Serbia
Marko Bohanec	Jozef Stefan Institute, Slovenia
Thomas Bournaris	Aristotle University of Thessaloniki, Greece
Lynne Butel	University of Plymouth, UK
Ana Paula Cabral	Federal University of Pernambuco, Brazil
Olivier Cailloux	University of Paris Dauphine, France
Guy Camilleri	Toulouse III University/IRIT, France
Hing Kai Chan	University of Nottingham, Ningbo Campus, UK/China
Alok Choudhary	Loughborough University, UK
João Paulo Costa	University of Coimbra, Portugal
Csaba Csaki	University College Cork, Ireland
Fátima Dargam	SimTech Simulation Technology/ILTC, Austria
Adiel Teixeira de Almeida	Federal University of Pernambuco, Brazil
Jorge Freire de Souza	Engineering University of Porto, Portugal
Pavlos Delias	Eastern Macedonia and Thrace Institute of Technology, Greece
Boris Delibašić	University of Belgrade, Serbia

Alex Duffy	University of Strathclyde, UK
Sean Eom	Southeast Missouri State University, USA
Lucia Fanini	IMBBC-HCMR, Greece
Gabriela Florescu	National Institute for Research and Development in Informatics, Romania
Xin James He	Fairfield University, USA
Jorge Hernández	University of Liverpool, UK
Jean-Marie Jacquet	University of Namur, Belgium
Uchitha Jayawickrama	Staffordshire University, UK
José Maria Moreno Jimenez	Zaragoza University, Spain
Milosz Kadzinski	Poznan University of Technology, Poland
Daouda Kamissoko	University of Toulouse, France
Gregory Kersten	Concordia University, Canada
Marc Kilgour	Wilfrid Laurier University, Canada
Kathrin Kirchner	Berlin School of Economics and Law, Germany
Giorgos Kotoulas	IMBBC-HCMR, Greece
Pheobe Koundouri	Athens University of Economics and Business, Greece
Priscila Lima	Federal University of Rio de Janeiro, Brazil
Isabelle Linden	University of Namur, Belgium
Shaofeng Liu	University of Plymouth, UK
João Lourenço	Universidade de Lisboa, Portugal
Michail Madas	University of Macedonia, Greece
Vasiliki Manthou	University of Macedonia, Greece
Jan Mares	University of Chemical Technology, Czech Republic
Bertrand Mareschal	Université Libre de Bruxelles, Belgium
Nikolaos Matsatsinis	Technical University of Crete, Greece
Brice Mayag	University of Paris Dauphine, France
Patrick Meyer	IMT-Atlantique, France
Evaggelos Mitsakis	HIT, Greece
Maria Morfoulaki	HIT, Greece
Festus Oderanti	University of Plymouth, UK
Alexandru-Liviu Olteanu	IMT-Atlantique, France
Petraq Papajorgji	Canadian Institute of Technology, Albania
Nikolaos Papandroulakis	IMBBC-HCMR, Greece
Jason Papathanasiou	University of Macedonia, Greece
Dobrila Petrovic	Coventry University, UK
Gloria Philipps-Wren	Loyola University Maryland, USA
François Pinet	Cemagref/Irstea, France
Nikolaos Ploskas	University of Macedonia, Greece
Ana Respício	University of Lisbon, Portugal
Rita Ribeiro	UNINOVA – CA3, Portugal
Alexander Smirnov	Russian Academy of Sciences, Russia
Athanasios Spyridakos	Piraeus University of Applied Sciences, Greece
Frantisek Sudzina	Aalborg University, Denmark

Theodore Tarnanidis	University of Macedonia, Greece
Stefanos Tsiaras	Aristotle University of Thessaloniki, Greece
Nikos Tsotsolas	University of Piraeus, Greece
Alexis Tsoukias	University of Paris Dauphine, France
Yvonne van der Toorn	Tilburg University, The Netherlands
Wim Vanhoof	University of Namur, Belgium
Leonilde Varela	University of Minho, Portugal
Konstantinos Vergidis	University of Macedonia, Greece
Rudolf Vetschera	University of Vienna, Austria
Maro Vlachopoulou	University of Macedonia, Greece
Isaak Vryzidis	University of Piraeus, Greece
Andy Wong	University of Strathclyde, UK
Lai Xu	Bournemouth University, UK
Pascale Zaraté	IRIT/Toulouse University, France

Steering Committee – EWG-DSS Coordination Board

Pascale Zaraté	IRIT/Toulouse 1 Capitole University, France
Fátima Dargam	SimTech Simulation Technology, Austria
Shaofeng Liu	University of Plymouth, UK
Boris Delibašić	University of Belgrade, Serbia
Isabelle Linden	University of Namur, Belgium
Jason Papathanasiou	University of Macedonia, Greece
Pavlos Delias	Eastern Macedonia and Thrace Institute of Technology, Greece

Local Organizing Team

Jason Papathanasiou	University of Macedonia, Greece
Nikolaos Ploskas	University of Macedonia, Greece
Theodore Tarnanidis	University of Macedonia, Greece
Georgios Tsaples	University of Macedonia, Greece
Panagiota Digkoglou	University of Macedonia, Greece
Georgios Kotoulas	Hellenic Centre for Marine Research, Heraklion, Greece
Panagiotis Adamidis	Alexander Technological Educational Institute of Thessaloniki, Greece
Lucia Fanini	IMBBC-HCMR, Greece

Sponsors

EWG DSS DECISION SUPPORT SYSTEMS	Working Group on Decision Support Systems (https://ewgdss.wordpress.com)
EURO THE ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIETIES	Association of European Operational Research Societies (www.euro-online.org)

Institutional Sponsors

	University of Toulouse, France (http://www.univ-tlse1.fr/)
	IRIT Institut de Research en Informatique de Toulouse, France (http://www.irit.fr/)
	SimTech Simulation Technology, Austria (http://www.SimTechnology.com)
	Graduate School of Management, Faculty of Business, University of Plymouth, UK (http://www.plymouth.ac.uk/)
	Faculty of Organisational Sciences, University of Belgrade, Serbia (http://www.fon.bg.ac.rs/eng/)
	University of Namur, Belgium (http://www.unamur.be/)
	University of Macedonia, Department of Business Administration, Thessaloniki, Greece (http://www.uom.gr/index.php?newlang=eng)
	Eastern Macedonia and Thrace Institute of Technology, Greece (http://www.teikav.edu.gr/index.php?lang=en)
	Hellenic Centre for Marine Research, Greece (http://www.hcmr.gr/en/)
	Center for Research & Technology, Greece (https://www.certh.gr/root.en.aspx)

(*Continued*)

	Alexander Technological Educational Institute of Thessaloniki, Greece (https://www.teithe.gr)

Industrial Sponsors

	REACH Consultancy (http://www.reach-consultancy.at/)
	Deloitte (https://www2.deloitte.com/)
	Springer (www.springer.com)
	Lumina Decision Systems (www.lumina.com)
	ExtendSim Power Tools for Simulation (http://www.extendsim.com)
	Paramount Decisions (https://paramountdecisions.com/)
	1000 Minds (https://www.1000minds.com/)

Contents

Decision Support Systems for a Sustainable Society

SIRPSS - Sustainable Industrial Site Redevelopment Planning Support System

Tong Wang(✉), Qi Han, and Bauke de Vries

Eindhoven University of Technology, 5612AZ Eindhoven, The Netherlands
tong.wang.0821@gmail.com

Abstract. Abandoned industrial sites should be redeveloped in a sustainable way to reduce waste land. Planning support systems (PSSs) can help in the early planning phase by presenting possible outcomes of various redevelopment scenarios to the planners. This paper presents a PSS to support sustainable industrial site redevelopment decision based on past experiences. Past experiences are recorded as cases in the system. Case-based reasoning (CBR) is applied. For sites needed to be redeveloped, similar redeveloped cases are retrieved from the case library. Strategies that the previous cases have applied are reviewed and the new site is designed based on the comparison study of new case and past experiences, together with the input from planners. Sustainability is dynamically evaluated on site and regional level. A showcase is presented for the system illustration.

Keywords: Planning support system · Case-based reasoning · Sustainability Redevelop strategy · Past experience

1 Introduction

1.1 Problem

Abandoned industrial land is a global concern. Local governments in China try to provide more industrial land to attract investment and compete with their neighboring cities [1]. As a result, during the period from 2001 to 2005, 8.78 million m^2 of industrial land is in need of redevelopment [2]. In the Netherlands, redundantly large supply of very cheap industrial sites causes similar problems [3, 4]. According to the Dutch national industrial site database IBIS, 1,052 industrial sites, equaling 28% of all industrial areas contain signs of obsolescence in 2012 [5].

The Netherlands is a densely populated country, with 501 people per square kilometers in 2014 [6]. It is important to reduce the amount of obsolete industrial sites for sustainable development. More and more attention is put on sustainable industrial sites redevelopment [7]. The new Spatial and Planning Act [8] is enforced in 2008 and it is expected to be integrated into the new Environment and Planning Act [9] in 2018. These new acts aim to simplify the procedures to speed up the decision-making process. This makes industrial site redevelopment process easier than before. In 2012, 13% of the identified obsolete places are planned for redevelopment, which amounts to a total area of 11.582 million m^2.

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 3–14, 2018.
https://doi.org/10.1007/978-3-319-90315-6_1

Planners can choose to redevelop an industrial site in many ways, but which way is the most suitable and why? Each region is different and therefore attention should be paid on each region's specific characteristics.

Decision support tools have been adopted in promoting industrial site redevelopment process. Specifically for selecting suitable industrial sites for redevelopment, Geographic Information System (GIS) is used by integrating various data in a geographical environment to prioritize the most suitable sites based on users' preferences [10]. Redevelopment outcomes are also assessed for different regions based on social, economic and environmental attributes [11–13]. However, redevelopment impacts on surrounding land use compositions are not yet addressed. The insight of land use transitions on surrounding land uses can help to evaluate possible future scenarios and help to find future suitable redevelopment sites. A system which can help analyze industrial land redevelopment process and understand its impacts on neighboring land uses after the transitions for each region is needed.

Moreover, to guide planners for the future land use design for a specific industrial site, it is beneficial to learn from past experience to avoid unnecessary failure. Land use development and redevelopment is an expensive and long term task. Using past experience can help reduce costs and risks. For one thing, both positive and negative experiences can be learned. For another, looking into the past can serve as a starting point for innovative designs for the future. In this paper, past experiences are recorded as cases in the planning support system. Cases here represent industrial sites redeveloped or under redevelopment processes. While in another example of such decision support tool proposed by Delias [14] to support building permit outsourcing decision, cases are then building permit application activities.

Past experience (cases) should be constructed in a database where users can consult the system to find similar past cases and use those experience to design and to evaluate the impacts for future design tasks, given the past case outcomes. For this purpose, Case-based Reasoning (CBR) approach can be applied. Indexing and retrieving past cases are of importance for such case library. Unique case representation structure should be designed to include the important attributes for describing each case and further matching past cases with the new case.

Sustainability evaluation should be included in the system as well. It should be not only on site level or on regional level since industrial sites may compete with each other in one region. As a result, sustainability indicators from both site and regional level should be used to compare different scenarios.

To conclude, abandoned industrial sites should be transformed in a sustainable way for a region. The impacts of such transition should be evaluated comprehensively. Computer aided tools like planning support systems could be designed to help this process by providing generic analyzing, designing and evaluating functions with the help of past experience (cases). Specific functions of this planning support system include selecting suitable sites for redevelopment, finding similar past experiences from the case library and designing new site redevelopment strategies and evaluating the impacts of such strategies for urban redevelopment process.

1.2 Literature Review

PSS. Planning support systems for strategic spatial planning, water management, transport networks and tourisms planning combined with scenario analysis are common examples [15]. For specific industrial land issues, not many references can be found except for industrial site selection, evaluation and risk management for contaminated sites [16, 17]. There are attempts for planning sustainable industrial area and modeling industrial land redevelopment collaboration process [18]. However, the former focuses on the evaluation of different scenarios and the latter is lack of spatial characteristics of land use planning.

Case-Based Reasoning. Case-based reasoning (CBR) provides approaches to solve problematic situations by finding similar cases from the past. A case is a contextualized piece of knowledge representing an experience that teaches a fundamental lesson to archive the goal of the reasoner, which normally contains a problem description and the solution to the problem [19].

The CBR approach has been formalized into four steps: retrieve, reuse, revise and retain as stated in [20]. In most cases, CBR is mainly used as a case retrieval system [19]. This is due to the fact that the revise process is not easy to automate.

Complex urban planning tasks require policy makers to discuss and adapt possible solutions, rather than relying solely on an automatically determined solution from similar cases. Therefore, in this research, focus is put on the retrieval process. In this process, similarity assessment is especially emphasized because of its importance for finding relevant cases [21–23]. To determine the similarity between cases, it is important to describe the cases in the same "language" - case representation structure which determines the way in which problems and solutions are encoded.

As stated by Wierzbicki [24], it is needed to show a kind of objectivity for situations where the decision may have impacts on many other people. Even though full objectivity is not achievable, rational objective ranking is desirable, with the help of available data and past references (cases). For industrial site redevelopment planning support which influences many parties involved, it is necessary to reach consensus on redevelopment directions among decision makers. In our paper, this type of consensus is referred as planning preferences for policy makers from each region. Different contexts have different planning preferences. These preferences can be elicited by analyzing regional zoning documents in which consensus are recorded and reflected by the frequency of words used. More details are presented on how to obtain these regional specific preferences [25]. To make it clearer for urban planners, these preferences are referred as visions, as commonly called in the zoning documents.

Meanwhile, only cases that have similar demographic and spatial characteristics should be considered as possible references for a new planning case [26, 27]. Quantifiable geospatial and demographic characteristics are instrumental for policy makers to make and evaluate planning decisions. As stated, not only the qualitative visions should be considered, but also quantitative attributes should be included to describe cases. The visions guide the redevelopment direction for a region while the quantitative attributes serve as input to the design and evaluation of specific strategies, and to find the most suitable redevelopment sites if needed.

1.3 Objectives

This project is to propose a planning support system to support sustainable industrial site redevelopment decision making. Past experience is stored, retrieved and learned for the new tasks. Land use simulation model is constructed based on the past land use dynamics. Important attributes for industrial site redevelopment design and sustainability are identified and evaluated.

2 Methodology

2.1 PSS Functions

Figure 1 presents the five modules of the proposed system. Data integration module combines regional and site level data. Site selection module chooses the most suitable sites for redevelopment based on regional visions and specific attribute requirements. Case library finds the most similar cases to the newly selected case. Their used strategies for redevelopment and the impacts of such redevelopment for the site and the region are presented. Based on the past experience, a possible redevelopment strategy for the new site can be chosen by the users. The determined strategy is then imported into land use simulation module. This module updates the land use situation dynamically and facilitates sustainability evaluation. Visualization module presents charts, numbers and maps.

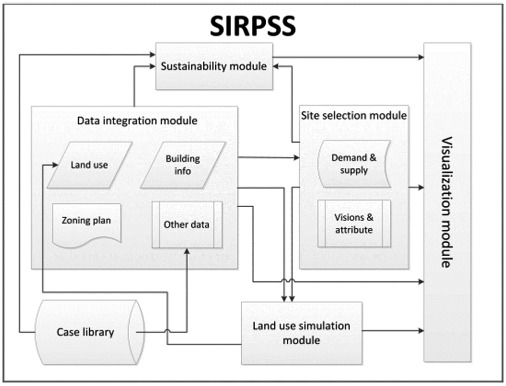


Fig. 1. SIRPSS modules

2.2 Research Flow

To explain the whole process, a research flow chart is presented in Fig. 2. Based on future users' requirements for a site and the regional visions from planners, suitable sites for redevelopment can be found based on the suitability calculation in GIS environment which integrates demographic data, building information and other GIS data. According to the same principles, similar redeveloped cases can be retrieved. Not only visions should be used to find similar past cases, but also other attribute values should be considered such as each site's accessibility to infrastructure, surrounding situations, population density and so on. After all, future users might have very specific requirements to one site like the soil cannot be heavily polluted. The visions help us to find the similar cases for the first step, and the following indicator or attribute values help us to find the most similar cases. The zoning documents for similar redeveloped cases can help compare the past experience with the current site at hand. Their similarities and differences can be identified to design the most suitable scenario for the new sites. The new design is imported into land use change simulation model so that future land use situation can be presented. The past experience from zoning documents and the new simulated land use situation for newly designed industrial site together help to evaluate possible outcomes of the new redeveloped site, both on site level and on regional level. This round can run several times for different scenarios and they can be compared.

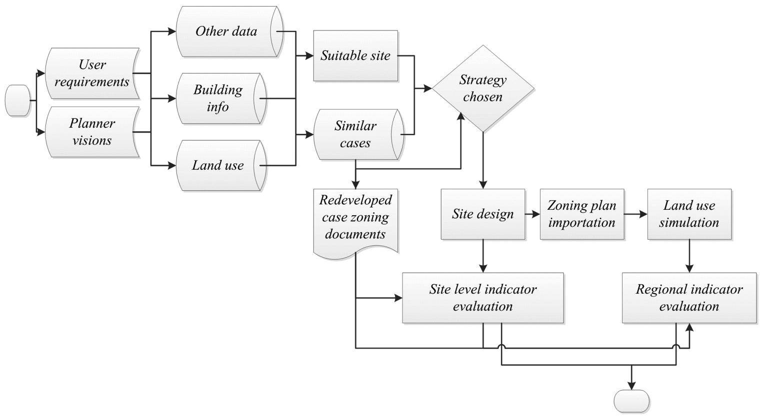


Fig. 2. Research flow for SIRPSS

3 Case Study

3.1 North Brabant Region

The North Brabant region has been chosen as the case study. Many industrial sites in this region have been redeveloped because of the limited land resources.

3.2 Data Collection

GIS data include land use maps [28], zoning maps [29], IBIS [5] and BAG [30] data. They are found from DANS, Niewskaart Nederland, North Brabant databank and PDOK [31] respectively.

Other economic, environmental and demographic data are mainly from CBS and national geography databank. SBI codes are the activities codes for Dutch industry [32].

They present the functions for each company in a building. SBI codes are used to integrate building information into GIS data so that more detailed description of past cases and the new site can be achieved.

3.3 Case Library

To find the redeveloped industrial sites in North Brabant region and construct such case library, a framework is presented [33]. In this paper, visions are extracted from zoning plan documents for the changed industrial sites in the North Brabant region using natural language processing tools. Besides the qualitative visions for redevelopment, quantifiable attributes values are also calculated based on the data available. This adds possibilities for evaluating redevelopment impacts analysis via comparing the measures of attributes value after and before redevelopment changes. The case representation structure is presented in detail in another paper [25]. Figure 3 presents the structure. Each case is represented by combining its transition form, regional redevelopment visions and quantifiable attribute values on site and regional level which can reflect redevelopment impacts.

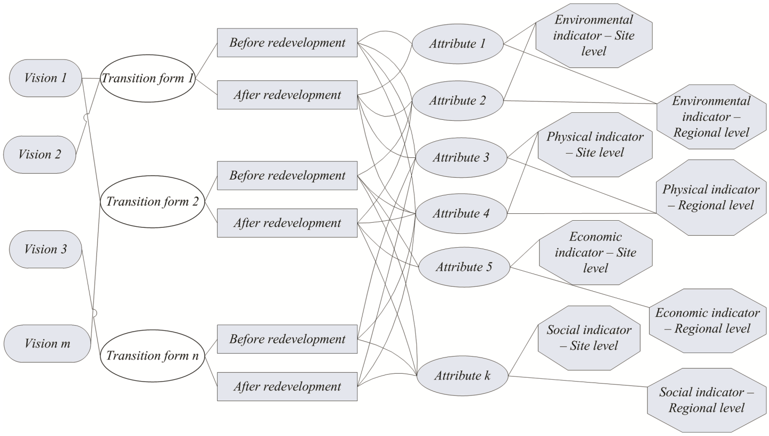


Fig. 3. Case representation structure

3.4 Suitable Redevelopment Cases

Visions. To showcase the PSS, the assumed vision for redevelopment is reuse of existing buildings and smooth transition; as a result, we have given high weights to total floor areas and light industry activities for each abandoned site in the region. As a result, sites with the most abandoned buildings and with light industry are selected. This is automatically executed in our PSS. Users can then change weights for quantitative attributes and then the similarity is automatically updated. More attributes can be added later based on regional development visions. A web tool can be found at http://sirpss.com/cases/.

3.5 Similar Case Retrieval

The most similar cases are found as Lichttoren + Victoriapark Eindhoven and Drie Hoefijzers based on the weight settings for current visions. Users can click on these two cases, detailed information is listed based on common structures of zoning plan documents, as shown in Table 1. Attributes are also included to show the dynamic temporal change such as population density before and after the redevelopment. More concrete information beside the calculated attribute values can also be consulted from the related zoning documents.

Table 1. Similar cases detailed information

Case	Drie Hoefijzers	Lichttoren + Victoriapark Eindhoven
Location	Nearby Station Breda	Eindhoven center
Size (hectares)	8	4.74
Plan use	Mixed	Business center, office, housing
Surrounding	A bigger development zone	City center
Noise	Optimal acoustic design	Exemption given for higher values
Soil	Needs to be remediated	Heavy metal contaminated
Energy usage	Goal: Energy neutral in 2015	A heat pump system; Renewable energy
Road design	A new residential street on west	Connection to highway A2/E25
Expected outcome	120 houses, southern part connects to the city with public spaces; northern separated from southern	Hotel: 15,500 m^2, Office: 3,650 m^2, Commercial: 4,250 m^2, Parking (276 spaces): 8,000 m^2, 128 lofts
Visions	Monumental buildings, Historical image, Sustainable energy, Public participation, Soil sanitation, Quality of life, Acoustic design, Mixed function, Archaeological values, Improve accessibility, Reuse existing buildings, Smooth transition	Industrial image, Historical image, High quality urban environment, Railway, Durability, Flexibility, Mixed function, Reuse existing buildings, Social housing, Expats, Monumental buildings, Green area, Service facility, Office, Improve quality, Livability, Smooth transition, High-rise buildings, Connection, Stay, Flexibility
Population density before	5,343 per km^2	5,555 per km^2
Residents before	1,450	990
Population density after	7,807 per km^2	5,496 per km^2
Residents after	2,130	1,080

3.6 Target Case Design

The situation for the target case is listed below in Table 2. Planners can use this information to compare the current case with past experience and design future uses.

Table 2. Target case detailed information including scenario design

Case	Target Case
Location	North of Bergen op Zoom, center of Halsteren
Size (hectares)	24.6
Previous use	Manufacture of concrete products; Manufacture of motorcycles; Renting of consumer goods; Shops selling water sports goods; Taxi operation; Sale and repair of passenger cars and light motor vehicles; Notaries; Social clubs; Petrol stations; Wholesale of computers; Wholesale of heating and cooling equipment; Sale and repair of passenger cars and light vehicles; Mortgage, bank and savings bank agencies
Plan use (hectares)	Mixed use, 9.3 for housing, 10 for public facilities, 5.3 for light industry
Surrounding	Near a bus stop (on the west), 1.5 km to train station at south, industrial area (north) with houses and green area (west)
Distance to city center	4.3 km to center of Bergen op zoom, center of Halsteren
Noise, Air quality, Odor	Near industrial area, noise should be reduced
Soil	Needs to be cleaned for concrete factory
Nature	Transition zones between the west and site for nature protection
Public transport	More bus lines

We have constructed a mixed use area where on the southeast it is transformed into housing since the southern area of the site is mainly residential land. North of the site is maintained to be industrial area since the northern side stays as a big industrial park. To the west side, there is green area so the west side of the site is transformed into public facilities to facilitate the new site and also provide smooth transition between the side and the surrounding green area. The scenario is presented in Fig. 4. Detailed scenario designs are implemented in ArcMap for site level indicator calculation. To incorporate this new design into the regional sustainability analysis, this scenario is imported as zoning plans into the land use change simulation module so that future land use situation can be simulated using such models. Regional sustainability indicators regarding physical conditions are therefore predicted.

Fig. 4. Target site design (Color figure online)

Sustainability Evaluation

Site Level. We used two indicators to illustrate the idea of using SIRPSS to evaluation site level sustainability, namely distance to public services and residence numbers. After the transition, with the new public facilities, site distance to public services is reduced to 0 m, while before the transition the minimum distance to public services is 500 m. Based on the past experience, one case increases 680 residence for an area of 8 ha while the other 90 for 4.74 ha. We then assume that the new case can increase residence by 52 (averaged by each hectare from two cases) * 9.3 ha for residential area based on the scenario design; which equals to 484.

Regional Level. We present one regional land use composition change situation to illustrate the influences of industrial site redevelopment on the region. Figure 5 shows that recreational land relative presence is increased after industrial sites being transformed into housing in the neighborhood area from 100 m to 200 m. In other words, industrial sites have a negative influence on the appearance of recreation, compared with housing's influence on recreation in this range for the whole region. This reflects that housing and recreational land has some level of synergy effect in short distance. After 200 m, there is less recreational land after the transition than before. Other indicators besides physical indicators can also be calculated with available data.

4 Conclusions and Discussion

This paper presents a planning support system to help design sustainable industrial site redevelopment strategies. The core of the system is case based reasoning. A case library is constructed. Based on the new case at hand, visions and specific requirements are provided by the users to find similar cases from the regional library. The past experience including their used strategies and impacts of such redevelopment is imposed onto the new case at hand. As a result, scenarios are designed for the new case and the sustainability impacts are evaluated on site and regional level. This system can help planners

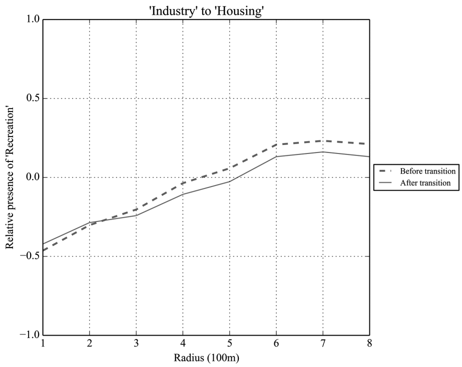


Fig. 5. Industrial land to housing impacts on regional land use compositions

and individual industrial land users to communicate different plans and evaluate the impacts on site and regional sustainability. It could also raise the enthusiasm for using the planning support systems. A region can even strengthen its competitiveness by offering such kind of system to allow individuals visualizing their own plans for the site.

More data should be collected to show the full functionality of sustainability evaluation. At this moment, only very limited indicators are calculated because of the limited data. Case library can be enriched using more data source and case representation structure can be further explored. Qualitative and quantitative attributes should be better combined to represent each case.

Workshops should be organized to ask policy makers opinions about this type of planning support systems applicability in their daily practice.

Acknowledgements. The corresponding author is financially supported by the China Scholarship Council under grant number [201207720002]. We would like to thank for their support.

References

1. Huang, Z., Du, X.: Strategic interaction in local governments industrial land supply: evidence from China. Urban Stud. **54**, 1328–1346 (2016)
2. Xie, J., Li, F.: Overview of the current situation on brownfield remediation and redevelopment in China. In: Sustainable Development - East Asia and Pacific Region Discussion Papers. World Bank, Washington, DC (2010)
3. Olden, H.: Remmende factoren bij het herstructureren van bedrijventerreinen. STOGO Onderzoek + Advies, Utrecht (2007)
4. Van Der Krabben, E., Buitelaar, E.: Industrial land and property markets: market processes, market institutions and market outcomes: the dutch case. Eur. Plan. Stud. **19**, 2127–2146 (2011)
5. Nijssen, C.E.G.E., Kremers, A.: IBIS werklocaties de stand van zaken in planning en uitgifte van werklocaties op 1 januari 2013 en de uitgifte in 2012. ARCADIS, Arnhem, The Netherlands (2013)
6. World DataBank: Population density (people per sq. km of land area). World DataBank (2014)

7. Louw, E., Bontekoning, Y.: Planning of industrial land in the Netherlands: its rationales and consequences. Tijdschr. Voor Econ. En Soc. Geogr. **98**, 121–129 (2007)
8. Summary National Policy Strategy for Infrastructure and Spatial Planning. Ministerie van Infrastructuur en Milieu, The Hague, the Netherlands (2011)
9. Environment and Planning Act. Ministerie van Infrastructuur en Milieu (2016)
10. Aktas, C.B., Bartholomew, P., Church, S.: Application of GIS to prioritize brownfield sites for green building construction based on LEED criteria. J. Urban Plan. Dev. **143**(3), 04017004 (2017). https://doi.org/10.1061/(ASCE)UP.1943-5444.0000387
11. Jigoria-Oprea, L., Popa, N.: Industrial brownfields: an unsolved problem in post-socialist cities. A comparison between two mono industrial cities: Resita (Romania) and Pancevo (Serbia). Urban Stud. **54**, 2719–2738 (2016)
12. Kiss, E.: Spatial impacts of post-socialist industrial transformation in the major Hungarian cities. Eur. Urban Reg. Stud. **11**, 81–87 (2004)
13. Ploegmakers, H., Beckers, P.: Evaluating urban regeneration: an assessment of the effectiveness of physical regeneration initiatives on run-down industrial sites in the Netherlands. Urban Stud. **52**, 2151–2169 (2015)
14. Delias, P.: A positive deviance approach to eliminate wastes in business processes: the case of a public organization. Ind. Manag. Data Syst. **117**, 1323–1339 (2017)
15. Geertman, S., Stillwell, J.C.H. (eds.): Planning Support Systems Best Practice and New Methods. Springer, Dordrecht (2009). https://doi.org/10.1007/978-1-4020-8952-7
16. Witlox, F., Timmermans, H.: MATISSE: a knowledge-based system for industrial site selection and evaluation. Comput. Environ. Urban Syst. **24**, 23–43 (2000)
17. Zabeo, A., Pizzol, L., Agostini, P., Critto, A., Giove, S., Marcomini, A.: Regional risk assessment for contaminated sites part 1: vulnerability assessment by multicriteria decision analysis. Environ. Int. **37**, 1295–1306 (2011)
18. Blokhuis, E., Snijders, C.C.P., Han, Q., Schaefer, W.F.: Conflicts and cooperation in brownfield redevelopment projects: application of conjoint analysis and game theory to model strategic decision making. J. Urban Plan. Dev. **138**, 195–205 (2012)
19. Biswas, S.K., Sinha, N., Purkayastha, B.: A review on fundamentals of case-based reasoning and its recent application in different domains. Int. J. Adv. Intell. Paradig. **6**, 235 (2014)
20. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. **7**, 39–59 (1994)
21. Cunningham, P.: A taxonomy of similarity mechanisms for case-based reasoning. IEEE Trans. Knowl. Data Eng. **21**, 1532–1543 (2009)
22. Lopez De Mantaras, R., Mcsherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K., Keane, M., Aamodt, A., Watson, I.: Retrieval, reuse, revision and retention in case-based reasoning. Knowl. Eng. Rev. **20**, 215 (2005)
23. Wess, S., Althoff, K.-D., Derwand, G.: Using *k*-d trees to improve the retrieval step in case-based reasoning. In: Wess, S., Althoff, K.-D., Richter, M.M. (eds.) EWCBR 1993. LNCS, vol. 837, pp. 167–181. Springer, Heidelberg (1994). https://doi.org/10.1007/3-540-58330-0_85
24. Wierzbicki, A.P.: The need for and possible methods of objective ranking. In: Ehrgott, M., Figueira, J.R., Greco, S. (eds.) Trends in Multiple Criteria Decision Analysis, vol. 142, pp. 37–56. Springer, Boston (2010). https://doi.org/10.1007/978-1-4419-5904-1_2
25. Wang, T., Han, Q., de Vries, B.: Sustainable industrial site redevelopment planning support: a case-based reasoning approach. Int. J. E-Plan. Res. **6**, 39–53 (2017)
26. Du, Y., Wen, W., Cao, F., Ji, M.: A case-based reasoning approach for land use change prediction. Expert Syst. Appl. **37**, 5745–5750 (2010)

27. Li, X., Liu, X.: An extended cellular automaton using case-based reasoning for simulating urban development in a large complex region. Int. J. Geogr. Inf. Sci. **20**, 1109–1136 (2006)
28. Centraal Bureau voor de Statistiek, Kadaster: Bestand Bodemgebruik 2003 – BBG 2003 (2003). https://doi.org/10.17026/dans-z8n-fc2u
29. Documentatie bij het GIS bestand van de Nieuwe Kaart van Nederland (2010). www.nieuwekaart.nl
30. Kadastre: BAGViewer (2017). https://bagviewer.kadaster.nl/lvbag/bag-viewer/index.html#?geometry.x=160000&geometry.y=455000&zoomlevel=0
31. PDOK: PDOK Viewer (2016). http://pdokviewer.pdok.nl/
32. Centraal Bureau voor de Statistiek: Standaard Bedrijfsindeling 2008. Centraal Bureau voor de Statistiek (2017)
33. Wang, T., Jessurun, J., Han, Q., de Vries, B.: A model to construct case base for sustainable industrial area redevelopment. In: Proceedings of IAMG 2015, pp. 1264–1272. International Association for Mathematical Geology (IAMG), Freiberg (2015)

An Intelligent Multi-agent System Using Fuzzy Analytic Hierarchy Process and Axiomatic Design as a Decision Support Method for Refugee Settlement Siting

Maria Drakaki[1], Hacer Güner Gören[2](✉), and Panagiotis Tzionas[1]

[1] Department of Automation Engineering, Alexander Technological Educational Institute of Thessaloniki, 141, 574 00 Thessaloniki, Greece
{drakaki,ptzionas}@autom.teithe.gr

[2] Department of Industrial Engineering, Pamukkale University, Kinikli Campus, Denizli, Turkey
hgoren@pau.edu.tr

Abstract. Crises and disasters of recent years are complex and occasionally interrelated phenomena, which require complex decision making for effective humanitarian relief operations provision. Moreover, coordination is needed between different humanitarian actors as decision makers. Massive refugee and migrant arrivals in Greece since 2014, mainly a result of a refugee crisis, require complex humanitarian supply chain management and logistics operations. In this paper, refugee settlement site planning decision making process is addressed with an intelligent multi-agent system (MAS) modeling approach. The MAS uses two multi-criteria decision making (MCDM) methods: fuzzy analytic hierarchy process (FAHP) to determine the weights of criteria and fuzzy axiomatic design approach extended with risk factors (RFAD) to rank alternative sites for refugee settlement. The proposed method will be applied to evaluate currently operating temporary accommodation sites in Greece.

Keywords: Intelligent multi-agent system · Refugee settlement planning
Refugee crisis · Decision support method

1 Introduction

As global forced displacement continues to grow, sea arrivals of refugees and migrants in the Mediterranean in 2015 reached a record high of 1.015.078 people. 851.319 arrivals were recorded in Greece in the same year [1]. In an evolving environment, in 2016, 362.753 arrived in the Mediterranean, whereas the number of those arriving in Greece was 173.450. As of September 2017, 49.3010 refugees and migrants reside in Greece, distributed in both mainland (35.057 people) and the islands (14.244 people). The Greek government is gradually closing temporary sites to move refugees into housing. However, in 2017, 25.694 refugees and migrants arrived in Greece, yet not at a constant rate, since

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 15–25, 2018.
https://doi.org/10.1007/978-3-319-90315-6_2

just for the month of September 6.000 sea arrivals were recorded. As a result, reception centers have already been overcrowded operating at a potentially unsafe environment.

The impact of the refugee crisis in Greece and thus to Europe as well, has been shown at a first level with the massive population arrivals since 2014, and a death toll exceeding 14.000 people in the Mediterranean in the same time period, including an estimated 3.000 people in 2017 [1]. Coordinated humanitarian efforts to cope with the refugee crisis and provide relief to the affected population involves many actors as in any humanitarian relief operation. These actors include the refugees and migrants, Greek government, local communities, United Nations High Commissioner for Refugees (UNHCR), European Union (EU), national and international Non-Governmental Organizations (NGOs), UN, national and international organizations and institutions, private and civic sectors, and donors.

Sites in Greece include Reception and Identification Centers (RICs), transit sites and Emergency Response Units on the islands. Currently more than 13.000 refugees and migrants reside on the islands at official sites and other state run facilities, far exceeding the corresponding less than 9.000 people capacity. On mainland, sites include temporary sites and sites on-hold, where people reside at official sites, other state-run facilities, sites provided by NGOs and UNHCR, or are self-settled. Data on site profiles in Greece are provided by UNHCR [1].

Strategic refugee settlement design should aim at ensuring a dignified and secure environment that enables social, economic and environmental quality of life for the refugee community. In addition, spatial planning design that considers settlement as a node rather than an "island", which is part of the economic, social and physical adjacent local territories contributes positively to a sustainable settlement and long-term organizing efforts [2]. Settlement and shelter design should follow the principles and standards of the SPHERE project [3]. Site selection and planning for refugees should be selected based on a set of key criteria and taking into account potential risk factors. Failure to follow minimum standards could expose settlement community to further risks and create tensions. The minimum shelter and settlement standards according to the SPHERE project include a range of factors for strategic planning, settlement planning, covered living space, construction and environment. Besides the SPHERE project, a set of most critical factors that should guide site selection have been proposed by UNHCR [4], as well as in the literature [5].

Research in humanitarian supply chain management and logistics has grown especially since 2004, however quantitative approaches and methods have not been applied extensively [6]. Research focus on refugee settlement planning, design and implementation has been given among others on strategic objectives for settlement and shelter [2, 7], and reports on existing settlements status [8]. Guiding principles include protection, i.e. the right to adequate housing, age, gender and diversity, equity, accessibility, access, sustainability, community empowerment, appropriateness and reliability including access to infrastructure and transportation networks, and environment including mitigation to the impact of natural environment. Moreover, settlement design should take into account host population and the context of the potential location area, as well as any evolving crisis. Additionally, potential ethnic and cultural affinities between refugees and host communities should be considered as well as potential friction between refugee ethnic groups. The complexity of refugee settlement planning points to adaptation of a multi-criteria decision making (MCDM) method. Quantitative methods, such as multi-criteria optimization methods is a rapidly growing literature trend in humanitarian aid operations [9]. A quantitative approach on refugee camp siting has been proposed in Cetinkaya et al. [5]. The authors followed a three step approach to identify potential refugee camp sites. Firstly, they identified relevant geographic, social, infrastructural and risk related criteria. Then, they entered criteria into a geographic information system (GIS) to obtain alternative locations for refugee camps. Weight values of GIS layers were subsequently determined using fuzzy analytical hierarchy process (FAHP). Alternative locations were then determined. Finally, technique for order preference by similarity to ideal solution (TOPSIS) method was applied to rank alternative locations. Key critical factors contributing to site selection planning include risk criteria, such as environmental hazards or community population ethnic distribution. Among MCDM methods proposed in the literature, hierarchical fuzzy axiomatic design (HFAD) with risk factors (RFAD) extends HFAD to integrate risk factors in the methodology, whereas it differentiates from the approach of considering risk factors as separate criteria [10].

MAS is an effective tool to describe complex phenomena, stochasticity and uncertainty. In the distributed artificial intelligence (AI) field, an agent is an autonomous software system, a problem solver, with local decision making, capable of interacting with other agents using communication and cooperation in order to solve a complex problem when it is beyond its own capability [11]. A MAS is defined as a set of agents that interact to achieve their individual goals, when they have not enough knowledge and/or skills to achieve individually their objectives [12]. Multi-agent systems (MAS) were initially deployed to address the features of intelligent manufacturing systems with a focus on providing decision robustness with respect to disruptions and complex dynamic environments [12].

In this paper, an intelligent MAS that uses an MCDM approach based on fuzzy analytic hierarchy process (FAHP) and RFAD is proposed as a decision support method for refugee settlement planning in Greece. The MAS consists of four agents, namely a site planner agent (SPA), a FAHP with extent analysis agent, a supervisor agent, and an RFAD agent. The MAS achieves its goal, i.e. evaluation and ranking of alternatives, by distributing the implementation of the MCDM method to a FAHP agent and a RFAD agent. The SPA decides the set of site selection criteria, risk criteria and corresponding risk factors, whereas the supervisor agent controls agent interactions in order that the final MAS goal is met. The proposed method will be applied to evaluate accommodation sites currently used in Greece. The MAS is implemented with JADE (Java Agent Development Framework) [13].

The rest of the paper is organized as follows. The next Sect. 2 gives some background information regarding MAS and JADE. The details of the proposed decision support system are explained in the Sect. 3 and the Sect. 4 summarizes the findings and presents some future research directions.

2 MAS

Modularity and abstraction, present in MAS, are necessary tools to solve complex, real-life problems. Interaction is achieved with communication. Coordination, co-operation and negotiation are important agent communication characteristics. Coordination between agents in order to achieve system goals is necessary when agents have different capabilities or when different agents work on the same goal in order to complete their tasks more efficiently. Negotiation among cooperative agents is an effective coordination technique for complex problem solving [13].

A MAS is hosted on a dedicated software platform and organized according to predefined principles such as agent architecture, communication language, and message protocols. The most popular agent architectures are the reactive and belief desire intention (BDI or deliberative) architectures. BDI agents exhibit human-like behavior. Reactive agents act with a stimulus-response mechanism. Reactive agents acquire intelligence as a result of their emergent behavior [13]. They do not have a model of the environment, the global system behavior is a result of their interactions, and thus, simplicity, robustness and fault tolerance are their key properties. A main disadvantage of the reactive agent architecture is that the prediction of global system behavior is not an easy task, i.e. it is a trial-and-error exercise based on experimentation since individual agents acquire local information only.

Multi-agent system (MAS) has been widely used in different applications such as diagnostics [14], energy management strategy [15] and dynamic real time rescheduling and learning [16]. In recent years, MAS has also been applied to problems in humanitarian logistics [17].

The most important properties of an agent for many application areas, including manufacturing systems, are autonomy, intelligence, adaptation, and co-operation [12], whereas intelligence can be increased through learning. Intelligent systems exhibit among others autonomous decision making, emergent behavior, adaptation, cost effectiveness and high quality.

Among agent concepts deployed in industrial applications, knowledge sharing and distributed learning are applied to industrial distributed diagnostics, whereas complex problems and dynamic real-world environments have been efficiently modeled with multi-agent simulation. Distributed problem using MAS solving seems appropriate for complex problems, as such problems ideally need more computational power. Decisions in MAS depend on individual agent local decision making as well as on negotiation and coordination mechanisms between agents.

Applications of MAS in decision support systems can be found in literature. MAS modeling was used for a decision support tool by Boutkhoum et al. [18], in which an integrated analysis method was based on MCDA combined with OLAP system to deal with complex decision making situations.

JADE is a widely used middleware for the development of FIPA compliant (Foundation for Intelligent Physical Agents) agent frameworks. It provides the infrastructure on top of which agent logic can be developed [13]. JADE includes the Java classes required to develop agents and the containers, i.e. the Java processes that support the run-time and the services which support agent existence and execution.

The JADE agent platform is the set of all active containers. Agent communication is compliant with FIPA and based on asynchronous message passing. The message structure format is compliant with the widely used FIPA-ACL (asynchronous communication language) with predefined interaction protocols (IPs).

3 Description of the Proposed Decision Support Method

3.1 The Elements of the Proposed Decision Support Method

The proposed decision support method is based on an intelligent MAS implemented with JADE reactive architecture. Agents use an MCDM method based on FAHP and RFAD. They communicate and interact using FIPA ACL. Each agent is created by defining a class that extends the jade agent class. The agents are created by defining classes that extend the jade agent class.

3.2 The MAS

The proposed MAS consists of four agents, namely a site planner agent, SPA, a RFAD agent, RFADA, a supervisor agent, SA and a FAHP agent, FAHPA. Agents interact using the FIPA request IP. Table 1 shows the agents of the presented MAS.

***The Site Planner Agent* (*SPA*)**
The SPA consults a database that contains site selection criteria collected from a range of data sources including UNHCR guidelines, the SPHERE project and literature. After discussions with all stakeholders, including government and host communities, as well as engineers, decides on the set of site selection criteria, i.e. main criteria and sub-criteria, as well as risk factors. Table 2 shows the minimum standards and a range of key critical factors and risk criteria to be considered when planning camps. The site planner receives the final ranking of alternatives from the coordinator agent and depending on the results, it approves it or initiates a new decision making cycle.

Table 1. MAS agents and their respective goals.

Agent	Goal
Site planner agent (SPA)	To decide the set of criteria and risk factors for site selection. To approve or disapprove the ranking of the alternatives and site selection
Supervisor agent (SA)	To coordinate and control agent interactions
FAHP agent (FAHPA)	To calculate priority weights of all criteria and sub-criteria, using FAHP
RFAD agent (RFADA)	To make the ranking of alternatives and site selection, using RFAD

Table 2. Minimum standards, key factors, other literature found factors and risk criteria for refugee site selection.

Minimum standards	Description
Covered living area	3, 5 sqm per person
Camp settlement size	Average camp area per person (30 sqm)
Fire safety	30 m of a firebreak every 300 m
Topography, soil condition, drainage, gradient for camp site	Topography of land should permit easy drainage. Land preferably covered with grass, with gradient ideally equal to 2 to 4% for good drainage
Drainage	Appropriate drainage should be put in place
Key factors	
Location-security: distance from international border and conflict zones	A distance of 50 km from international border is preferable
Location-security, distance from environmentally protected area	Site should be located within 1 day's walk from the protected area
Supportive factors: accessibility to national services, such as health, education, markets and towns	Site's proximity to main services, e.g. health services, markets and towns
Supportive factors: accessibility to roadway	Access to road infrastructure should be reliable
Potential beneficiaries: the numbers	High population numbers could lead to increased tension and protection threats to vulnerable groups, as well as health risks
Potential beneficiaries: the type of PoC	Relative proportions of ethnic groups as well as age and gender of PoC contribute to protection risks
Potential beneficiaries: the length of stay	Service provision over a long period is challenging. It could result in stress, tensions and social conflict between ethnic groups. It could also distort local economies
Location: land mines	Land mines pose threats to the PoC
Basic characteristics of the land: land use and land right	Public land preferable
Basic characteristics of the land: elevation	Elevation is related to flood risks. Sites should be located above flood level
Basic characteristics of the land: soil condition	Soil erosion results in environmental degradation. Rocky, impermeable soil should be avoided.
Basic characteristics of the land: water availability	Site located close to a water source
Basic characteristics of the land: sanitation	Sufficient sanitation facilities should be provided. Access to sanitation is a human right
Basic characteristics of the land: climatic conditions	Sites should be located away from regions that experience extreme climatic conditions

(*continued*)

Table 2. (*continued*)

Minimum standards	Description
Basic characteristics of the land: vegetation	It is preferable that land permits cultivation and vegetable gardens
Supportive factors: harvesting wood	Wood is a shelter material that could be collected if permitted by the local authorities
Supportive factors: availability of electricity	Electricity should be provided
Supportive factors: nearby villages or communities	Links with local communities should be established
Supportive factors: proximity to agriculture or income generation activities	Income generation is a basic livelihood standard
Supportive factors: availability of UN agencies, NGOs, operational partners, humanitarian groups	Liaises with UN agencies, NGOs and humanitarian groups are preferable
Additional factors from literature [5] - proximity to poverty	Sites should be located close to welfare areas for resource availability
Additional factors from literature [5] -distance from tourism attractiveness	Sites should be away from tourism attractions
Risks	
Location-security, environmental risks	Risks related to natural hazards, such as earthquakes, high winds, fire risks, flooding, landslide
Location-security, health risks	Locations that present health risks (e.g. malaria) should be avoided
Security and protection, high population density	Increased exposure to health risks, tension, and protection threats to vulnerable groups
Security and protection, the type of PoC	Exposure of PoC to protection threats, tensions between ethnic groups

The FAHP Agent (FAHPA)

The FAHPA uses the FAHP method with extent analysis to prioritize the weights of criteria. A fuzzy comparison matrix contains the pairwise comparisons related to determining the weights of criteria.

AHP [19] is a process for developing a numerical score to rank each decision alternative based on how well each alternative meets the decision maker's criteria. It allows users to assess the relative weight of multiple criteria or multiple options against given criteria in an intuitive manner. In some decision problems, all data are not available so in order to deal with such a decision problem, fuzziness should be added to the solution approaches. Chang's Extent Analysis [20] method on Fuzzy AHP is used in this paper. Let be

$$\text{Goal set} = U = \{u_1, u_2, \ldots., u_n\}$$
$$\text{Object set} = X = \{x_1, x_2, \ldots., x_n\}$$

$$\text{Extent analysis values} = \left(M_{gi}^1, M_{gi}^2, \ldots.., M_{gi}^m\right)$$

Triangular fuzzy numbers with the parameters $l, m, u = M_{gi}^{j} (j = 1, 2, \ldots., m)$
The steps of Chang's extent analysis [20];

1. The value of fuzzy synthetic extent degree with respect to the i-th object is defined as:

$$S_i = \sum_{j=1}^{m} M_{gi}^{j} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{n} M_{gi}^{j} \right]^{-1} \quad (1)$$

To obtain $\sum_{j=1}^{m} M_{gi}^{j}$, the fuzzy addition operation is performed for m extent analysis values such that:

$$\sum_{j=1}^{m} M_{gi}^{j} = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j \right) \quad (2)$$

To obtain the $\left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{gi}^{j} \right]^{-1}$, perform the fuzzy addition operation of M_{gi}^{j} $(j = 1, 2, \ldots., m)$ values such that:

$$\sum_{i=1}^{n} \sum_{j=i}^{m} M_{gi}^{j} = \left(\sum_{i=1}^{n} l_i, \sum_{i=1}^{n} m_i, \sum_{i=1}^{n} u_i \right) \quad (3)$$

The inverse of the vector in Eq. 3 is calculated as follows:

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{gi}^{j} \right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} u_i}, \frac{1}{\sum_{i=1}^{n} m_i}, \frac{1}{\sum_{i=1}^{n} l_i} \right) \quad (4)$$

2. The degree of possibility of two triangular fuzzy numbers M_1 and M_2 is defined as:

$$V(\tilde{M}_2 \geq \tilde{M}_1) =_{y \geq x} \sup[\min(\mu_{\tilde{M}_1}(x), \mu_{\tilde{M}_2}(y))] \quad (5)$$

The equation can be equivalently expressed as follows:
where d is the ordinate of the highest intersection point D between M_1 and M_2

$$V(\tilde{M}_2 \geq \tilde{M}_1) = hgt\,(\tilde{M}_1 \cap \tilde{M}_2) = \mu_{M_2}(d) \quad (6)$$

3. The possibility degree that a convex fuzzy number is greater than k convex fuzzy numbers M_i (i=1, 2,…., k) can be defined as;

$$\begin{aligned}
&V(M \geq M_1, M_2, \ldots..., M_K) = V[(M \geq M_1) \quad ve(M \geq M_2) \qquad ve\ldots(M \geq M_k)] \\
&= \min V(M \geq M_i) \\
&1,\ 2,\ 3,\ \ldots.,\ k \\
&\forall k = 1,\ 2,\ 3,\ \ldots\ldots,\ k \\
&k \neq i \\
&d'(A_i) = \min V(S_i \geq S_k) \text{ for} \qquad k = 1, 2, \ldots..., n.
\end{aligned} \tag{7}$$

where $A_i(i = 1,\ 2,\ 3, \ldots..., n)$ are n elements.

The normalized weight vectors via normalization are as follows:

$$W = (d(A_1), d(A_2), \ldots\ldots,\ d(A_n)) \tag{8}$$

where W is a non-fuzzy number. This gives the priority weights of one alternative over another.

The RFAD Agent (RFADA)

In order to consider risk factors in the decision problem, HFAD with risk factors (RFAD) has been utilized in this study. The HFAD has been extended to include risk factors by Gören and Kulak [10] and RFAD has been proposed. The RFAD integrates risk factors in its methodology. The RFADA uses the RFAD method to evaluate and rank alternatives to make a final site selection decision. The agent uses the priority weights calculated by the FHAPA and the risk factors decided by the SPA for the final ranking.

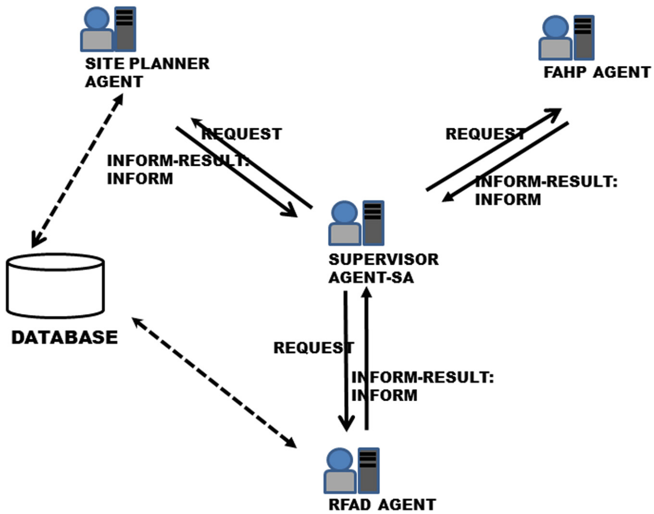


Fig. 1. Agent interaction diagram in the presented MAS.

3.3 The Agent Interaction Protocol

SPA sends the selected criteria and risk factors to the SA with a request communication act. SA sends the selected criteria to the FAHPA with a request communication act for evaluation of the criteria. FAHPA sends the priority weights to the SA with an inform-result communication act. SA sends the priority weights to the RFADA and the risk factors with a request communication act. The RFADA sends the final site location decision and ranking to the SA with an inform-result communication act. The SA sends the final decision to the SPA with an inform-result communication act. If the SPA agrees on the site selection, the process has completed. Otherwise, it can restart the whole process. Figure 1 depicts the agent interaction diagram.

4 Conclusions

In this paper, an intelligent MAS is presented in decision making of refugee camp siting. Refugee camp siting involves a complex decision making process involving different decision makers. The multi-agent system models and solves the problem by distributing related tasks to different agents representing the decision makers, until an optimal or near optimal solution is obtained. Agents in the MAS use a hybrid MCDM method based on FAHP and RFAD. They are implemented with JADE and communicate and interact using FIPA ACL. Each agent is created by defining a class that extends the jade agent class. A coordinator agent supervises and controls agent interaction and communicates with the site planner agent, initially to receive the list of criteria, and finally to deliver the ranking of alternatives for approval. Risk criteria are integrated as risk factors in the land criteria by the RFAD agent. The learning agents, i.e. FAHPA and RFADA, acquire knowledge by employing FAHP and RFAD respectively. The final normalized weights of criteria are calculated by the FAHP agent and transferred to the coordinator agent. The RFAD agent forwards the data to the RFAD agent that calculates the ranking of the alternatives. The procedure can be repeated until the site planner agent approves the results. The proposed method will be applied to evaluate four alternative currently used accommodation sites in Greece. Different MCDM methods in the MAS might be used as future research directions.

References

1. UNHCR data portal (2017). https://data2.unhcr.org. Accessed 17 Nov 2017
2. Moore, B.: Refugee settlements and sustainable planning. Forced Migr. Rev. **55**, 5–7 (2017)
3. SPHERE Sphere Project, Sphere Handbook: Humanitarian Charter and Minimum Standards in Disaster Response (2011). http://www.ifrc.org/docs/idrl/I1027EN.pdf. Accessed 20 Nov 2017
4. UNHCR emergency handbook: UNHCR Handbook for Emergencies. https://www.unicef.org/emerg/files/UNHCR_handbook.pdf. Accessed 20 Nov 2017
5. Çetinkaya, C., Özceylan, E., Erbaş, M., Kabak, M.: GIS-based fuzzy MCDA approach for siting refugee camp: a case study for southeastern Turkey. Int. J. Disaster Risk Reduction **18**, 218–231 (2016)

6. Beamon, B.M., Kotleba, S.A.: Inventory modelling for complex emergencies in humanitarian relief operations. Int. J. Logistics Res. Appl. **9**(1), 1–18 (2006)
7. Terne, M., Karlsson, J., Gustafsson, C.: The diversity of data needed to drive design. Forced Migr. Rev. **55**, 25–26 (2017)
8. Wain, J.F.: Shelter for refugees arriving in Greece. Forced Migr. Rev. **55**, 20–22 (2017). 2015–17
9. Gutjahr, W.J., Nolz, P.C.: Multicriteria optimization in humanitarian aid. Eur. J. Oper. Res. **252**, 351–366 (2016)
10. Gören, H.G., Kulak, O.: A new fuzzy multi-criteria decision making approach: extended hierarchical fuzzy axiomatic design approach with risk factors. In: Dargam, F., et al. (eds.) Decision Support Systems III - Impact of Decision Support Systems for Global Environments. LNBIP, vol. 184, pp. 141–156. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-319-11364-7_13
11. Wooldridge, M., Jennings, N.R.: Intelligent agents: theory and practice. Knowl. Eng. Rev. **10**, 115–152 (1995)
12. Leitao, P.: Agent-based distributed manufacturing control: a state-of-the-art survey. Eng. Appl. Artif. Intell. **22**, 979–991 (2009)
13. Bellifemine, F.L., Caire, G., Greenwood, D.: Developing Multi-agent Systems with JADE, vol. 7. Wiley, Hoboken (2007)
14. Drakaki, M., Karnavas, Y.L., Chasiotis, I.D., Tzionas, P.: An intelligent multi-agent system framework for fault diagnosis of squirrel-cage induction motor broken bars. In: Świątek, J., Borzemski, L., Wilimowska, Z. (eds.) ISAT 2017. AISC, vol. 656, pp. 80–89. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-67229-8_8
15. Dou, C.X., Wang, W.Q., Hao, D.W., Li, X.B.: MAS-based solution to energy management strategy of distributed generation system. Electr. Power Energy Syst. **69**, 354–366 (2015)
16. Narayanaswami, S., Rangaraj, N.: A MAS architecture for dynamic, realtime rescheduling and learning applied to railway transportation. Expert Syst. Appl. Int. J. Arch. **42**(5), 2638–2656 (2015)
17. Edrissi, A., Poorzahedy, H., Nassiri, H., Nourinejad, M.: A multi-agent optimization formulation of earthquake disaster prevention and management. Eur. J. Oper. Res. **229**, 261–275 (2013)
18. Boutkhoum, O., Hanine, M., Agouti, T., Tikniouine, A.: Multi-agent based modeling using multi-criteria decision analysis and OLAP system for decision support problems. World Acad. Sci. Eng. Technol. Int. J. Comput. Inf. Eng. **9**(12), 2553–2560 (2015)
19. Saaty, T.L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
20. Chang, D.Y.: Applications of extent analysis method on fuzzy AHP. Eur. J. Oper. Res. **95**(3), 649–655 (1996)

Towards an Intelligent Integrated System for Urban Planning Using GIS and Cloud Computing

Boudjemaa Khelifa[1], Mohamed Ridda Laouar[1(✉)], and Sean Eom[2]

[1] Laboratory of Mathematics, Informatics and Systems (LAMIS), University of Larbi Tebessi, 12002 Tebessa, Algeria
khelifa_b@yahoo.fr, ridda_laouar@yahoo.fr
[2] Southeast Missouri State University, Cape Girardeau, USA
sbeom@semo.edu

Abstract. During the last few years, urban field has become more and more complex, as much as the growth of new technologies, new problems occurred. Therefore, urban territory planning organizations look ahead to find out sustainable alternatives, to meet human urban needs and to reach a high level of efficiency with the best employment of available resources. This paper proposes an Intelligent Integrated System using Geographic Information System (GIS) and Cloud Computing technology. The projected approach is based on: (i) the expand of the local GIS, by adding required information; (ii) the prediction of urban needs for a local scope, via analyzing GIS information; and (iii) the support of decision-makers so as to fit suggested urban projects to appropriate areas, respecting constraints. This paper includes a new architecture combining the main components of the proposed system.

Keywords: Urban planning · Intelligent decision support system · GIS
Cloud computing · Sustainable development

1 Introduction

Recently, the world has witnessed deep and rapid transformations in various areas of life. Where the most affected elements concern the urban planning fields. Which makes difficult the control of such situation using common systems. Therefore, it is necessary to find out sustainable and effective alternatives to meet human needs and to raise urban planning challenges. These alternatives coincide with the achievement of the Millennium Development Goals (MDGs) social priorities about poverty, hunger, disease, unmet schooling, gender inequality, and environmental degradation.

This paper, introduces an Intelligent Integrated System using Geographic Information System (GIS) and Cloud Computing technology, towards a strategic urban planning process. By taking profit of the up-to-date techniques and technological platforms, considered as solutions for both structured and unstructured urban planning problems.

The proposed approach gives a technical support to cover the most information around urban planning fields, including means for data acquisition, processing,

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 26–37, 2018.
https://doi.org/10.1007/978-3-319-90315-6_3

analyzing and visualization. The main function we focus-on is to extract easily the urban needs about a local scope, by predicting requirements for urban development projects, to support decision-makers (DM) to fit available urban projects to appropriate locations.

In order to attain an efficient development, and to master the process of complex urban planning, we look ahead to establish an Intelligent Decision Support System (IDSS) which deals with different components belonging to various fields of urban planning, defined as interacting coherent units that can affect and be affected by each other. Every unit concerns a set of social priorities with measurable objectives.

1.1 Urban Planning Challenges

According to the studies of successive civilizations, urban planning is dating back too far. It convoyed the human evolution, and its effects prove human ability to adapt to imminent fears. To maintain the existence and to improve the future of the community [1], implicated actors: planners, decision-makers, politicians, economists, stakeholders, and mainly urban authorities must challenge and overcome urban problems, which are depending on energy consumption, natural resources preservation, and environmental pollution.

Urban planners usually encounter the following problems: urban growth and land-use [2–4], infrastructures and facilities (water, food, energy and supplies) [5–8], transportation, communication networks [9], housing [10], education [11], healthcare [12], disaster preparedness [13], and environmental issues (air, noise and pollution) [14, 15]. They act as well as to reach ultimate policy and sustainable goals, which requires adapted evaluation methods and tools to make decisions [16].

In urban planning cases, the process of decision making often leads to manipulate conflicting objective functions, which makes the task more complex [17], whereas urban phenomena engenders a huge volume of information and highly depends on the appropriate location. Consequently, this arise the need to obtain only reliable, complete and readily available data as well as operational tools. Therefore, urban planners and professionals should get profit of advanced analysis modeling tools, using intelligent decision support system (IDSS), geographic information system (GIS) and Cloud Computing.

1.2 Intelligent Decision Support System

Intelligent decision support system (IDSS) is the combination of Artificial intelligence (AI) and decision support systems (DSSs) providing new computing assistance to operate on complex environment [18–20]. In order to simplify the decision-making process under uncertain, incomplete and vague situations, IDSS works well using sophisticated computer programing tools.

New decisions depend on previous system outcomes. In such cases, IDSSs reacts similarly as human reasoning to point out the best alternatives to reach optimized suitable results throw many leading techniques, like knowledge-based systems, artificial neural network systems, intelligent agent-assisted decision support systems, fuzzy systems, evolutionary computing and genetic algorithm based decision support systems (GA-based DSS), rough set based DSS, data mining, and process mining.

1.3 Geographic Information System

Geographic information system (GIS) technology enables to visualize, question, and analyze data to understand relationships, patterns, and trends in different fields. The vulgarization of GIS use, helps significantly the emergence of powerful applications for urban planning to local government operations. It facilitates the efficient and portable spatial data storage, updating and processing [21]. Nowadays, GIS technology has completely dominated human life. It becomes accessible to all organizations and even to the public. Although in the past, it was reserved to military authorities, specific departments and research centers. Modern communication technologies, such as internet, smartphones and cloud computing, are contributing to empower GIS, and make it widely operational. Currently, departments and government agencies are introducing this system more and more, in different fields such as public safety, public works and utilities, and environmental protection and monitoring.

1.4 Cloud Computing

Cloud computing, is an internet metaphor based on the concept of virtualization, it applies the principle of computing as utility, and provides shared data and processing resources for users on demand. According to The US National Institute of Standards and Technology (NIST), cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interactions [22].

The NIST describes principal service models of cloud computing: Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS), and Cloud Infrastructure as a Service (IaaS). Also, the cloud computing deployment models: Private cloud, Public cloud, Hybrid cloud and Community cloud.

This paper is organized as below, an introduction resuming the purpose work involving the main elements related to the topic, then a Sect. 2 detailing the current research context against exposed problems. Some related works are cited in the Sect. 3, whereas the forth section focuses on the suggested solution, and finally a last concluding the whole work.

2 Context of the Research

Through organisms, agencies and public administration, the government adopts strategies to contribute in the local development, by providing necessary resources with the formulation of ways and means, to satisfy citizen's needs and well serve the general welfare. Usually development programs concern different points such as: (1) employment and wealth creation, (2) enhancement of living conditions, (3) economic stability and food security, (4) health care and prevention, (5) education system, (6) protection against disaster, (7) preservation of environment and sustainable development (8) guarantee the national sovereignty.

To master the above points, it is necessary to manipulate up-to-date information; therefore, public and private actors spend exhausting efforts, to collect and to analyze required data provided from several sources. The issued data is often characterized by enormous volume, redundancy and uncertainty, which makes the process of extracting usable data a big challenge. In spite of the high dependency between urban elements, the coordination between different actors remains feeble and causes the waste of time and money.

The aim of this work is to propose an intelligent integrated system for urban planning, to assist planners and decision-makers, who share the same objectives in strategic, tactical and operational levels, in order to achieve the process of local development regarding the sustainability's goals. Therefore, it is important to benefit from cloud computing facilities combined with geographic information system (GIS) analyzing capabilities, to increase the quality of decision making process for urban planning (see Fig. 1).

Fig. 1. Use of GIS and cloud computing for urban planning improvement.

The proposed system provides technical support to cover information concerning urban planning fields, including means for data acquisition, processing, analyzing, and visualization. The main functions we focus on are to extract easily the urban needs about a local scope, and to support decision-makers to fit available urban projects to the appropriate areas (see Fig. 2).

Local development covers adjacent areas, belonging to the same local authority, where each area has potentials and requirements in different urban fields. In order to satisfy citizen needs, authorities give directives to all implicated actors regarding to their domains. According to available resources, those actors aim to accomplish urban projects, but it is difficult to do the asked tasks easily at time, because of lack involvement of different parts, especially those extra to related development field. An overview on neighboring areas is required to facilitate the allocation of urban project in a rational, fair and just way.

For example, for areas A_1, …, A_n, with a deficiency in the number of classrooms, respectively x_1, …, x_n, we tend to provide classrooms as much as necessary. When the

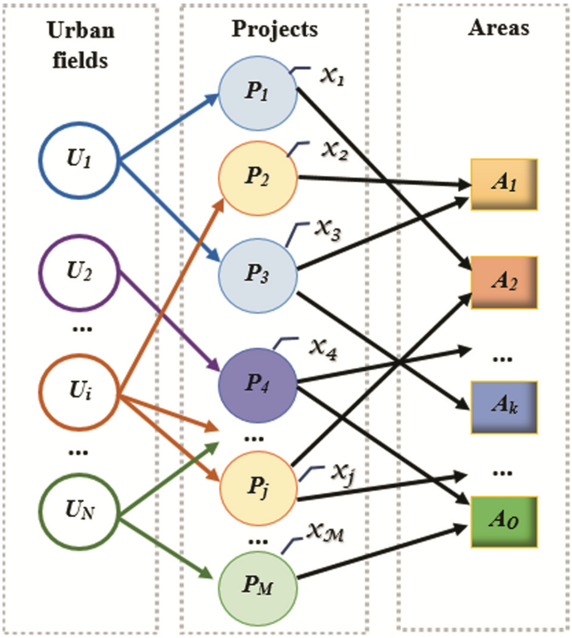


Fig. 2. Sample for complex urban planning problem, the fit of available urban projects to the appropriate areas with respect to urban fields.

budget suffice to build only a limited number of schools, it is better to allocate them to the areas with maximum deficiency in order to avoid more school transport costs. This may be also depends on provided facilities, environment and safety factors.

Another example, allocating investment projects to appropriate areas should respect some factors, such as the vocation of area (agriculture-oriented, industry-oriented, commerce-oriented, or touristic-oriented), field saturation, raw material availability, and projects integrity. To proceed, this process takes much time to obtain agreements and permissions because of shortage of information about locations that will host those projects.

Planners and decision-makers work together for public interest. They plan and manage urban projects continuously to find better solutions, considering all objectives. The process of decision making often leads to manipulate conflicting objective functions, which makes the task more and more complex, besides that urban phenomena engenders a huge volume of information and highly depends upon the appropriate location.

In this context, urban projects planning is a selection of a multitude development projects to promote a specific area according to the given needs. The projects are modeled as elementary units. For example, in schooling the elementary project is the class, as for healthcare we associate a hospital bed to an elementary unit. Therefore, it is necessary to provide enough project's units, as much as the area needs, but it is not always the case, especially when we met various problems and limited resources.

3 Related Works

Too many works in the field of urban planning are performed, using geographic information system and cloud computing, among them the following:

In the research of Coutinho-Rodrigues et al. in 2011, they developed a decision support system to provide decision aid in urban infrastructure planning. They noticed that the visualization of available alternatives on maps provides a value-added for decision support process in urban infrastructure evaluation problems [7].

While the research of Khan et al. in 2012, they advocate the use of cloud technologies to support the information, communication and decision-making needs of a wide variety of stakeholders in the complex business of the management of urban and regional development [23].

In the research of Tu et al. in 2015, to overcome the problem of unacceptable logistics service delay caused by the unpredictable traffic accidents and traffic states in the metropolitan area. They developed a spatiotemporal decision support (STDS) framework to facilitate large scale logistics distribution in the metropolitan area [24].

Meanwhile, Chen et al. in 2015 implemented an Intelligent Analysis and Mining System for Urban Lighting, using data analysis, case-based reasoning, data mining and GIS. This system detects lighting problems, provides early warning alarm, and also supply constructive information for the urban development [25].

Finally, in 2016 a recent research of Azevedo et al. proposed an approach for integrating heterogeneous data located in various public organizations using GIS, decision making process in the context of minimizing damage caused by floods [26].

Due to limited processing resources, lack of information exchange and bad coordination between concerned actors by urban planning process, especially for those sharing the same concerns. Most of elaborated works in the field, as mentioned above often focused on homogenous projects, with limited set of criteria ignoring an essential factor, called the interdependency between urban field's components.

In spite of the successful spread of the new technologies of cloud computing, and its facilities to provide infinite processing resources and services, particularly public actors still not interested to take advantage of these technologies because of unawareness and fear about information security and national sovereignty.

In fact, scientists should find out adequate solutions to acquire user's confidence, showing that the supposed fears are negligible beside the great facilities offered by new technologies.

4 Proposed Solution

Through this work, we expect to improve the performance of IDSS with the integration of GIS, and Cloud Computing, in order to tackle complex urban planning problems for strategic management in local development.

Each local scope is characterized by its own urban data, considered as a whole entity. For that reason, urban planners and decision-makers should define adequate strategies

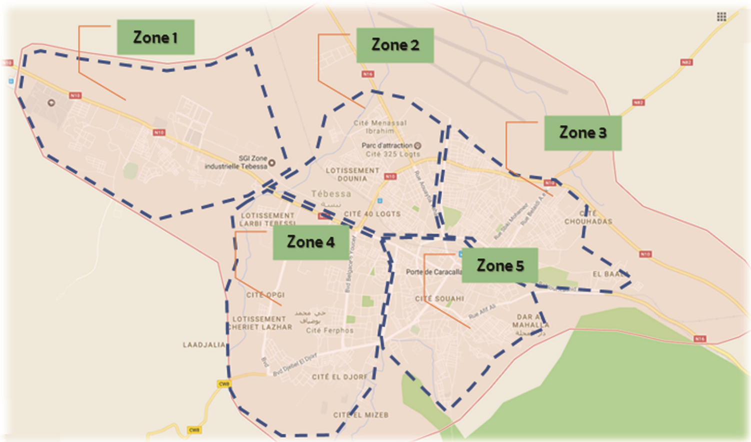


Fig. 3. Partitioning areas into small zones.

for zone partitioning, regarding geographic features, socio-economic factors and administrative partition plan (see Fig. 3).

The zoning concept contributes openly in urban planning tasks, but it couldn't express clearly all the urban needs in details. Therefore, it is obvious to conceive a suitable architecture, which highlights the layer's elements for each urban field. The

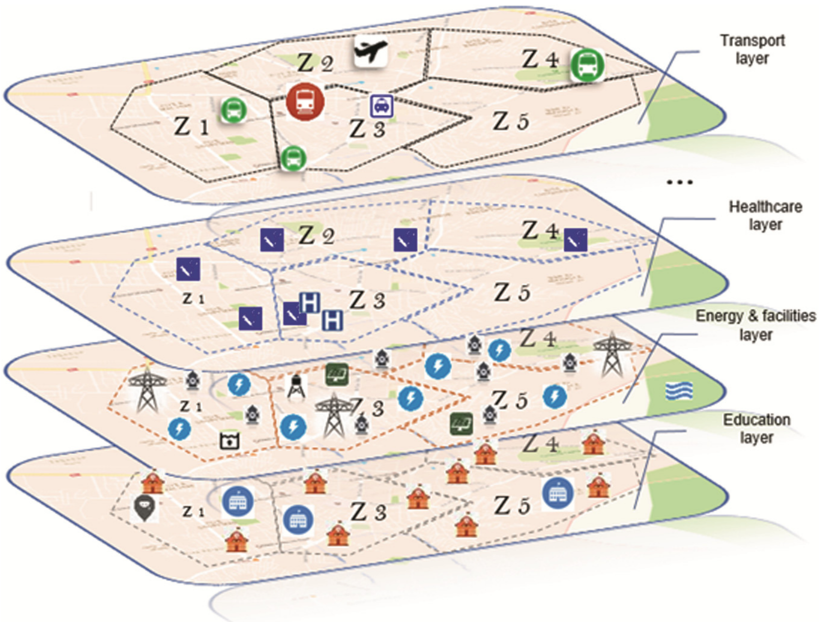


Fig. 4. Visualization of multilayered urban planning fields.

resulting multi-layered architecture must conserve the zone information as much as the details of each single urban field (see Fig. 4).

The projected system integrates necessary means and takes benefit from common GIS APIs and open data sources (e.g., Natural Earth Data[1], Esri Open Data[2], OpenStreetMap[3], and FAO GeoNetwork[4]).

Based on relevant linked data sources and services, this architecture represents clearly each urban field as one layer, which concerns one of the followings: health, energy, housing, education, industry, agriculture, religion, sports, transportation and tourism…etc.

According to the proposed approach, urban planners should follow the next steps:

- *The first step* consists of enriching the information system by providing necessary socioeconomic indicators respectively for local, regional and national levels. We should also include all available information (records and databases), to complete, upgrade and validate the information extracted from the linked sources (e.g. open WebGis).
- *The second step* is to apply IDSS techniques to analyze provided data, and to find out patterns about the distribution of basic infrastructures and resources over the scope, with regard to socioeconomic indicators. Through those patterns, the planners analyze the current situation and get a clear overview to promote the scope.

To predict the urban field's new projects, that match better to an appropriate area, an IDSS is required. It operates on available information about infrastructures with an associated catalogue, which provides exhaustive description of urban components grouped by categories. For example, by using the catalogue, a health infrastructure has its own proprieties, which distinguishes it from others. Therefore, we need more features such as type, category, accommodation capacity, principal vocation, coverage rate, power consumption, costs and others socioeconomic impacts.

- *In the last step,* planners and decision-makers are able to finalize the process of urban planning, based on predicted projects as mentioned above. Decision-makers rely on the IDSS to obtain the best combinations of projects that match urban goals respecting resources constraints. The goals are built on good practices in the urban field, which cover the whole system within layers and areas (see Fig. 4).

To ensure a better integrity for the urban system, we analyze horizontally the components of all contiguous areas field by field, and vertically area by area (see Fig. 4). This process leads to employ search techniques on massive data, which requires immense computation resources. Heuristic search techniques show its efficiency to solve this kind of problem, where new technologies provide necessary resources at any time and everywhere, to respond to current urban planning requirements.

[1] Natural Earth was built NACIS (North American Cartographic Information Society), and is free for use in any type of project, http://www.naturalearthdata.com/.

[2] Environmental Systems Research Institute, http://www.esri.com/.

[3] https://www.openstreetmap.org.

[4] Interactive Maps, GIS datasets, Satellite Imagery and Related Applications, http://www.fao.org/geonetwork/srv/en/main.home.

It is necessary to host the essential databases and services on the cloud, because the use of cloud computing and linked open source GIS data offers widely solicited resources with much profit and less cost, and consequently helps to achieve urban planning process.

The resulting architecture formed by Web GIS, community cloud, urban infrastructures catalogues, and multilayered GIS, surrounding urban planning projects, traces previous urban status, and gives an overview of coming projects to develop (see Fig. 5).

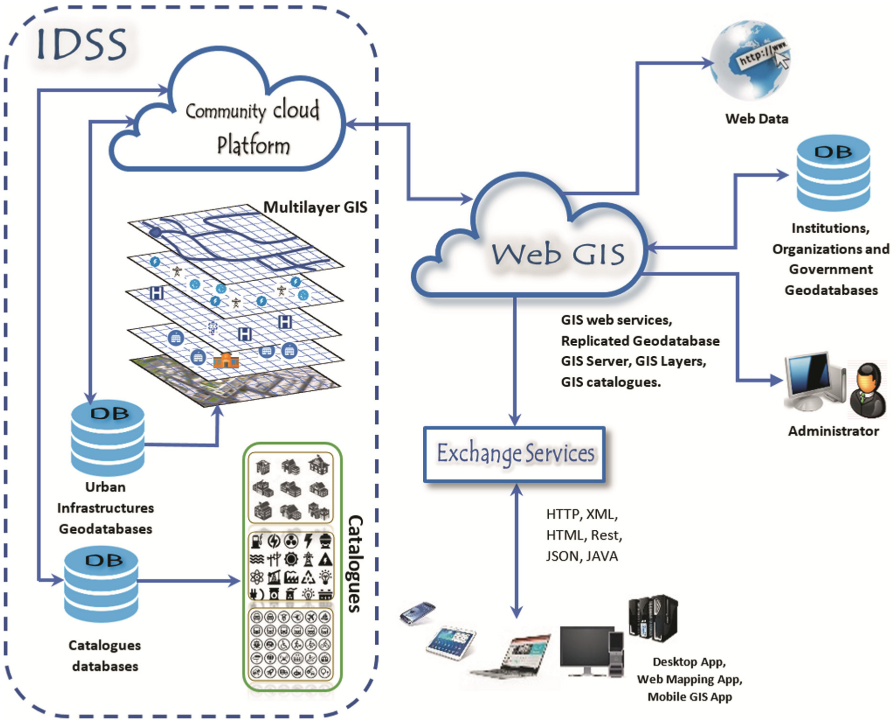


Fig. 5. Architecture of an IDSS integrating GIS and cloud computing, for complex urban planning problems.

To visualize cartographic data and linked open source GIS, we propose to incorporate specializing open source Application Programming Interfaces (APIs) such as: Quantum GIS (QGIS), Geographic Resources Analysis Support System (GRASS GIS) and Google GIS.

Pertinent urban planning data includes area's details (infrastructure labels, coordinates, adjacency, demographic and socioeconomic indicators, and area vocations… etc.) for each field. To exchange information, standardized methods for information transmission on the Internet should be used, such as eXtensible Markup Language (XML) and Javascript Object Notation (JSON). Those information should be coherent and exhaustive to make better decisions.

The deployment model for this IDSS (left side Fig. 5) is the community cloud, which is provisioned for exclusive use by a specific community of consumers from organizations that have shared same concerns.

For a specific area, urban field users have permission to operate directly on their own databases, and through Web GIS platform they access to community cloud data, and consequently planners and decision-makers achieve theirs tasks quickly and effectively by means of instantaneous information issued from different domains.

5 Conclusion and Outlook

The proposed approach supports decision-makers to accomplish the process of a sustainable development, in local level covering different urban fields. We have to extend the simple development cycle for urban planning, which is somehow closed on itself, and to recommend more interaction between urban components by exploiting the necessary means (IDSS techniques, SIG, Cloud Computing). A new architecture combining the main components of the proposed system offers a suitable geospatial data employment (representation, processing, visualization, sharing, and management).

Using the infrastructure description catalogue, the system power resides in the facility of extraction and enhancement concerning usable and valuable data. It provides a full technical support that covers mostly information around urban planning fields, which simplify the prediction of urban needs at a local scope. It relies on both linked open source GIS data and hosted community cloud data.

Our perspective is to achieve the implementation of the proposed architecture and to apply it on high dimensionality urban planning cases. Also, we are planning to automate the process of areas partitioning based on geographical data, administrative districts and socioeconomic indicators. This makes the approach more convivial for decision-makers to ensure efficient urban planning process.

Acknowledgments. The authors would like to thank the referees for their efforts to read and comment this manuscript.

References

1. Yin, J.: Urban Planning for dummies. Wiley, Hoboken (2012)
2. Dai, F., Lee, C., Zhang, X.: GIS-based geo-environmental evaluation for urban land-use planning: a case study. Eng. Geol. **61**(4), 257–271 (2001)
3. Al-shalabi, M., Billa, L., Pradhan, B., Mansor, S., Al-Sharif, A.A.: Modelling urban growth evolution and land-use changes using GIS based cellular automata and SLEUTH models: the case of sana'a metropolitan city, Yemen. Environ. Earth Sci. **70**(1), 425–437 (2013)
4. Abdullahi, S., Pradhan, B.: Application of GIS and RS in urban growth analysis and modeling. In: Pradhan, B. (ed.) Spatial Modeling and Assessment of Urban Form, pp. 297–311. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54217-1_13
5. Shen, L., Cheng, S., Gunson, A.J., Wan, H.: Urbanization, sustainability and the utilization of energy and mineral resources in China. Cities **22**(4), 287–302 (2005)

6. Doukas, H., Patlitzianas, K.D., Iatropoulos, K., Psarras, J.: Intelligent building energy management system using rule sets. Build. Environ. **42**(10), 3562–3569 (2007)
7. Coutinho-Rodrigues, J., Simâo, A., Antunes, C.H.: A GIS-based multicriteria spatial decision support system for planning urban infrastructures. Decis. Support Syst. **51**(3), 720–726 (2011)
8. Kabir, G., Sadiq, R., Tesfamariam, S.: A review of multi-criteria decision-making methods for infrastructure management. Struct. Infrastruct. Eng. **10**(9), 1176–1210 (2014)
9. Black, J.A., Paez, A., Suthanaya, P.A.: Sustainable urban transportation: performance indicators and some analytical approaches. J. Urban Plan. Dev. **128**(4), 184–209 (2002)
10. Türkoklu, H.: Residents' satisfaction of housing environments: the case of Istanbul, Turkey. Landsc. Urban Plan. **39**(1), 55–67 (1997)
11. Adomßent, M., Fischer, D., Godemann, J., Herzig, C., Otte, I., Rieckmann, M., Timm, J.: Emerging areas in research on higher education for sustainable development–management education, sustainable consumption and perspectives from central and eastern Europe. J. Clean. Prod. **62**, 1–7 (2014)
12. McLafferty, S.L.: GIS and health care. Annu. Rev. Public Health **24**(1), 25–42 (2003)
13. Shah Alam Khan, M.: Disaster preparedness for sustainable development in Bangladesh. Disaster Prev. Manag. Int. J. **17**(5), 662–671 (2008)
14. Andersson, E., McPhearson, T., Kremer, P., Gomez-Baggethun, E., Haase, D., Tuvendal, M., Wurster, D.: Scale and context dependence of ecosystem service providing units. Ecosyst. Serv. **12**, 157–164 (2015)
15. de Roo, G.: Integrating City Planning and Environmental Improvement: Practicable Strategies for Sustainable Urban Development. Routledge, Abingdon (2017)
16. Yaakup, A., Bakar, S.Z.A., Sulaiman, S.: Decision support system for urban sustainability planning in Malaysia. Malays. J. Environ. Manag. **10**(1), 101–117 (2009)
17. Torkamani, F., Fallah, S., Saadatmand, S.: How urban managers can use DSS to facilitate decision-making process: an application of fuzzy TOPSIS. J. Am. Sci. **8**(5), 162–173 (2012)
18. Sprague Jr., R.H.: A framework for the development of decision support systems. MIS Q. **4**, 1–26 (1980)
19. Turban, E., Aronson, J., Liang, T.P.: Decision Support Systems and Intelligent Systems, 7th edn. Pearson Prentice Hall, Upper Saddle River (2005)
20. Das, T.: Intelligent techniques in decision-making: a survey. Indian J. Sci. Technol. **9**(12) (2016)
21. Arampatzis, G., Kiranoudis, C.T., Scaloubacas, P., Assimacopoulos, D.: A GIS-based decision support system for planning urban transportation policies. Eur. J. Oper. Res. **152**(2), 465–475 (2004)
22. Mell, P., Grance, T., et al.: The NIST definition of cloud computing (2011)
23. Khan, Z., Ludlow, D., McClatchey, R., Anjum, A.: An architecture for integrated intelligence in urban management using cloud computing. J. Cloud Comput. Adv. Syst. Appl. **1**(1), 1 (2012)
24. Tu, W., Li, Q., Chang, X., Yue, Y., Zhu, J.: A spatio-temporal decision support framework for large scale logistics distribution in the metropolitan area. In: Harvey, F., Leung, Y. (eds.) Advances in Spatial Data Handling and Analysis. AGIS, pp. 193–206. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19950-4_12
25. Chen, G., Wang, E., Sun, X., Tang, Y.: An intelligent analysis and mining system for urban lighting information. Int. J. Smart Home **9**(3), 253–262 (2015)
26. Azevedo, P.C.N., Pinto, V.A., Bastos, G.S., Parreiras, F.S.: Using linked open data in geographical information systems. In: Grueau, C., Gustavo Rocha, J. (eds.) GISTAM 2015. CCIS, vol. 582, pp. 152–166. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-29589-3_10

27. Benatia, I., Laouar, M.R., Bendjenna, H., Eom, S.B.: Implementing a cloud-based decision support system in a private cloud: the infrastructure and the deployment process. IJDSST **8**(1), 25–42 (2016)
28. Benatia, I., Laouar, M.R., Eom, S.B., Bendjenna, H.: Incorporating the negotiation process in urban planning DSS. IJISSS **8**(2), 14–29 (2016)

An Ontology-Based Decision Support Framework for Personalized Quality of Life Recommendations

Marina Riga[1,2(✉)], Efstratios Kontopoulos[1], Kostas Karatzas[2], Stefanos Vrochidis[1], and Ioannis Kompatsiaris[1]

[1] CERTH-ITI, Information Technologies Institute, Thessaloniki, Greece
{mriga, skontopo, stefanos, ikom}@iti.gr
[2] ISAG-EI Group, Department of Mechanical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
{mriga, kostas}@isag.meng.auth.gr

Abstract. As urban atmospheric conditions are tightly connected to citizens' quality of life, the concept of efficient environmental decision support systems becomes highly relevant. However, the scale and heterogeneity of the involved data, together with the need for associating environmental information with physical reality, increase the complexity of the problem. In this work, we capitalize on the semantic expressiveness of ontologies to build a framework that uniformly covers all phases of the decision making process: from structuring and integration of data, to inference of new knowledge. We define a simplified ontology schema for representing the status of the environment and its impact on citizens' health and actions. We also implement a novel ontology- and rule-based reasoning mechanism for generating personalized recommendations, capable of treating differently individuals with diverse levels of vulnerability under poor air quality conditions. The overall framework is easily adaptable to new sources and needs.

Keywords: Personalized decision support · Ontology · OWL 2 SPIN · Air quality recommendations · User profiling

1 Introduction

In the environmental domain, it was early recognized that there is a positive correlation between the quality of the atmospheric environment and people's quality of life (QoL), acknowledging the fact that sensitive parts of the population suffer as atmospheric quality parameters become worse. Although air quality (AQ) experts could benefit from rapid advances in information technology and telecommunications, which enabled the efficient monitoring, analysis, transmission and sharing of scientific AQ data, it was only after the European Union's Directive 97/101/EC that environmental information should mandatorily become available and easily accessible to the public via official and direct communication channels.

As a first result, several urban air quality management and information systems emerged. However, the delivered content was usually limited to spatiotemporal air

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 38–51, 2018.
https://doi.org/10.1007/978-3-319-90315-6_4

quality observations, emission data, or AQ forecasts [1], with no sufficient explication of their meaning or impact to individuals' health. The provision of the aforementioned content in a way that can support end-users in decision-making activities is an integral part of the so called *environmental decision support system* (*EDSS*).

The way communication of AQ information is handled in EDSSs is important for their wider acceptance. Humans perceive environmental quality on the basis of their personal interests; they need to be informed if caused events have an impact on the daily living [2, 3]. Hence, systems that link the *location*, the *incident*, and the *individual* are of high interest for the end-users [4]. In this direction, relevant initiatives, such as PESCaDO[1], AirForU[2], Clean Air Nation[3] and Air Visual[4], demonstrate the added value of real-time, location-based AQ information and recommendation services. However, these applications do not handle user-profiles' differentiation, but produce general advice under poor AQ conditions that universally apply to sensitive people and not to the specific user who queries for decision support (DS).

Our motivation is to integrate the involved data and processes of an EDSS in a uniform, modular and user-profile centric framework. The effectiveness of this demanding task lies in the following subtasks: (i) to efficiently handle the heterogeneous and multifaceted nature of data, (ii) to adequately associate the provided experts' knowledge and rules for targeted, user-profile driven recommendation provision, and (iii) to facilitate scalability and reusability of the framework to third-party modules. In this context, we take advantage of the semantic expressiveness of ontologies to deal with the above issues. *Ontologies* are state-of-the-art Semantic Web technologies for structuring and semantically integrating heterogeneous content. Among other applications, they have been successfully adopted for covering individual parts of the decision making process [5]: (i) collecting, storing, and processing data, (ii) formulating the decision-making problem, (iii) reasoning over the data to reach decisions.

In contrast to existing ontology-based decision support systems (DSSs) that merely exploit the semantic web technologies in parts of the decision making process (see next section), our proposed approach demonstrates the extensive use of ontologies and semantic reasoning technologies by handling both the *static* (representation) and *dynamic* processes (realization, inference) of a DSS operation. The proposed schema comprises a set of ontological concepts and relations for semantically representing a primitive section of experts' knowledge and AQ dynamics, focusing on definitions of air pollutants' measurements, health risks, sensitive groups, and relevant user-profile driven recommendations. We implement a novel rule-based ontological reasoning mechanism for routing the problem of dynamic classification and new knowledge extraction to support personalized recommendation provision. To the best of our knowledge, no other DSS covers this multifaceted task as a whole through the adoption of ontologies. The proposed work comprises the operational EDSS of the hackAIR EU project [6].

---

[1] http://pescado-project.upf.edu/.

[2] http://newsroom.ucla.edu/releases/new-app-lets-you-check-air-quality-as-easily-as-checking-the-weather.

[3] http://www.greenpeace.org/india/Clean-Air-Nation/.

[4] https://airvisual.com/app.

2 Related Work

An EDSS literature review reveals various implementations, with respect to: (i) the application domain for DS (e.g. urban air quality management, extreme climate risks administration, etc.) and (ii) the technological approaches (e.g. computational intelligence methods, mathematical models, etc.). A detailed review of EDSS technologies, tools, and use cases is presented in [7, 8]. In most related work items, DS is addressed for administrative purposes, involving experts or regulatory authorities as the targeted end-users. Instead, our proposed framework aims to support QoL services for the general public, with an additional strategic difference against existing implementations, the process of "translating" AQ observations into user-profile driven recommendations, for personalized guidance in severe atmospheric conditions.

Considering ontologies in the environmental domain, numerous implementations exist covering the representation of abstract, general concepts (e.g. SWEET, a modular schema with 6,000 concepts in 200 ontologies describing concepts of physical and ecological phenomena, meteorological conditions, processes, activities [9]) as well as domain-specific, applied concepts (e.g. the PESCaDO ontology for personalized environmental DS [5], the EnvO ontology for the concise description of environmental features, materials and habitats [10], the AIR_POLLUTION_Onto ontology for air pollution analysis and control [11], etc.). Inspired by their expressiveness and adaptability, we build our relevant representations as described in the next section.

Ontologies have been proposed in EDSS for different tasks: in [12] for semantic search and easy access of structured environmental data; in [13] for integrating existing local databases of environmental data as part of the Linked Open Data cloud, enabling the linking of data in an established context and the dissemination of environmental information to the masses; in [14] for facilitating the process of selecting domestic solar hot water systems according to specific criteria, and in [15] for integrating heterogeneous content from multiple environmental sources. Despite the increasing deployment of ontology-based solutions in DSS, their potential is merely exploited, either for creating a structured representation of the domain of interest, or for supporting parts of the decision making process. With our proposed framework, we demonstrate the efficient use of ontologies and their supported technologies in all the basic components of the DSS.

Similarly to our approach, within the context of the PESCaDO EU project, ontologies were used as the backbone of the proposed EDSS, supporting all phases of the decision making process [16]; nevertheless, its rules along with the reasoning module are hardcoded in the source code, resulting in a highly inflexible approach. On the other hand, our implementation pushes the usage of ontologies one step further: both the domain knowledge and the experts' rules are developed at the ontology level, with the use of the OWL language and SPIN rules (see next section). The proposed DS framework operates as a stand-alone and uniformly developed module that can be easily adoptable by external sources, independently of their implemented technologies. It is also flexible and extensible, in terms of easy initialization of the different concepts, rules and recommendations even by non-ontology experts, by simply following the definitions of the proposed schema.

3 Proposed Approach

In computer science, an ontology is defined as a formal explicit specification of the terms and relations that describe a domain of discourse in a structured and semantically rich way [17]. The adoption of ontologies enables the understanding, sharing and reuse of information among different systems. Their capabilities fit perfectly to the task of describing and integrating heterogeneous content, and of dynamically inferring new knowledge, in a multidisciplinary field of study such as air quality. In this paper, we do not exhaustively represent the complete dynamics and facts or relevant association rules existing in the AQ domain; instead we demonstrate a practical paradigm of ontology use for real scenarios, conforming to related literature suggestions stating that citizens as end-users seek *personalized information services*, with *timely* or *in advance* AQ information provision, with respect to *their location* of interest [4].

The proposed approach handles all three basic components of a general DSS [18] through the use of ontologies: (i) the *data* component, i.e. the ontology-based schema developed for representing an excerpt of AQ domain experts' knowledge and of AQ information provision; (ii) the *model* component, which is an ontology-based and augmented with rules dynamic representation of principles that generate recommendations by combining different disciplines in the field of interest; and (iii) the *user-system interaction* component, which involves the communication of the user with the system. All of the above are more thoroughly described in the following subsections.

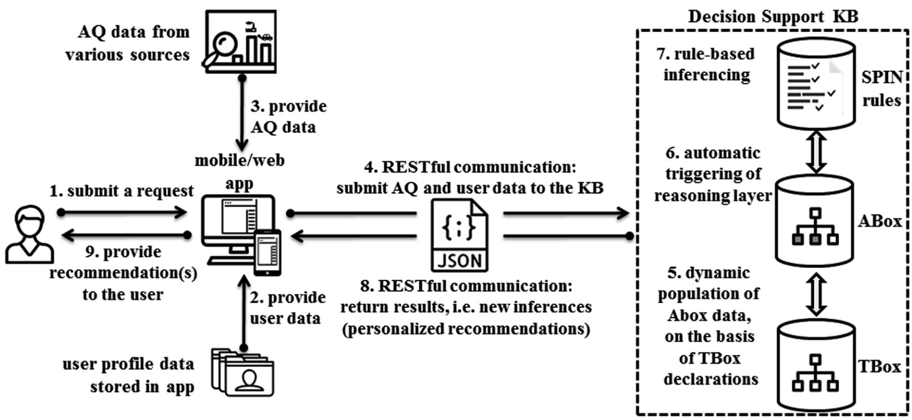


Fig. 1. Overall architecture and workflow of the proposed approach

In this context, we implemented a *3-layered* decision support knowledge base (KB) on the basis of three interconnected ontologies: (1) the *TBox*, i.e. the bottom layer that formalizes information relevant to the concepts of discourse (user profile, AQ measurements, requests and personalized recommendations) and their interrelations, (2) the *ABox*, i.e. the middle layer that is based on the schema from the TBox declarations and formalizes information relevant to membership/attribute assertions (actual users, observations, etc.), and (3) the *rules* layer, i.e. the upper layer that is based on both the aforementioned layers and formalizes the set of rules for reasoning (inference) different levels of interpretations. A general overview of the proposed ontology-based architecture and information workflow is presented in Fig. 1.

3.1 Domain Knowledge Modelling

The representational primitives utilized in ontologies are: *classes* referring to concepts or (abstract) entities that are assumed to exist in some domain of interest, *individuals*, which are instantiations (i.e. objects) of the classes, and *properties*, i.e. relationships that hold among objects. Their formal definition within the ontology typically carries information about concepts' meaning (semantics) and constraints that exist within the actual context of the domain of discourse.

To elicit the requirements that our ontology should satisfy, we followed the guidelines proposed by the *NeOn methodology* [19]. First, we defined the multifaceted *purpose of use* of our ontology, which includes the following goals: (i) to serve as an operational framework for the representation and orchestration of heterogeneous environmental, health, user profile-related data; (ii) to integrate the rules that govern air pollution and their impact to QoL, according to provided environmental and health experts' knowledge; and (iii) to support user-oriented recommendation services, with respect to personal health/user preferences (i.e. activities, daily routine, asthma, etc.) and to current AQ conditions for the location of interest.

Then, we identified the ontology's *intended users* as: (i) individuals, people with health sensitivities, those working or exercising outdoors, all those interested in receiving information about existing AQ conditions, so as to limit their exposure to hazardous conditions, or to increase their awareness about the impact of air pollution under special circumstances; (ii) public administrators or environmental experts, interested in receiving AQ information for professional reasons; (iii) technology experts, developers or ontology engineers, interested in adopting and expanding the ontology model for relevant DSSs and services.

The aforementioned definitions implicitly define the content and structure of the ontology. The latter is developed in OWL 2, a W3C standard ontology language [20]. Here, we present an excerpt of conceptualizations that are significant for adequately structuring the DS process following a user request. An indicative graph of the main classes and their relationships, developed in *Grafoo* [21], is presented in Fig. 2.

Class `Person` encapsulates those individuals' characteristics that are required for the reasoning process, specified on the basis of the available health-related advice and knowledge. We define distinct subclasses of class `Person`, with respect to the following parameters: (i) year of birth (`ChildPerson`, `ElderlyPerson` etc.), (ii) health sensitivity (`PregnantFemalePerson`, `SensitiveHealthPerson`), (iii) daily preferred activities (`SportsWalkingPerson`, `OutdoorJobPerson` etc.). Specific *ontology rules* (see next section) and *class expression axioms* (e.g. `EquivalentClasses (OutdoorJobPerson ObjectSomeValuesFrom (worksOutdoors "true"^^xsd:boolean)`)) declare the underlying semantics and restrictions.

In our proposed schema, a user may be classified as an instance of more than one types (e.g. `SensitiveHealthPerson` and `SportsWalkingPerson`), unless

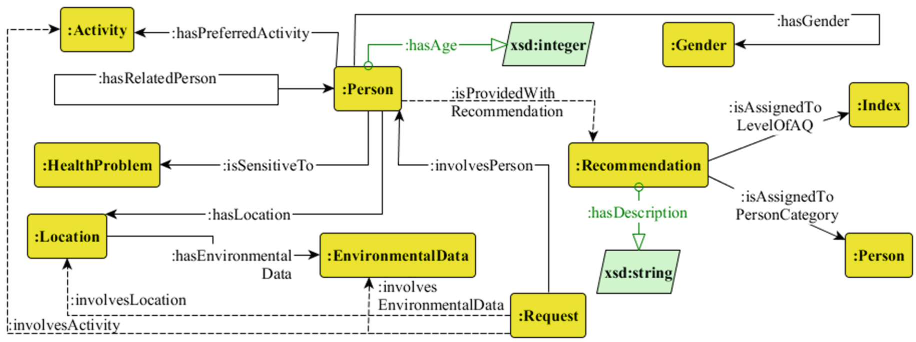


Fig. 2. An excerpt of classes and relations of the proposed ontology; properties with dashed line are inferred via rules.

such combinations are forbidden (e.g. `DisjointClasses (ElderlyPerson Child-Person)`). Moreover, a `Person` instance can be associated with other linked user profiles, enabling a *one-to-many* DS provision. An example case could be a mother with respiratory problems requesting personalized AQ recommendations for herself and her child, where separate recommendations will be produced from a single request, with respect to each profile's characteristics.

Class `Activity` represents instances of indoor/outdoor activities. On the basis of rules, recommendations to the users differ according to the nature or intensity of the activity; e.g. in case of harmful AQ conditions, activities like running or biking may lead to increased oxygen uptake and should be avoided or replaced with alternatives.

Class `HealthProblem` represents health problems (e.g. respiratory, cardiovascular, circulatory diseases) that the DSS takes into account when generating recommendations with respect to users' health sensitivities, under poor air quality conditions.

Class `Location` represents the location of interest (city, country or as coordinates). The more discretization introduced in the ontology, the more complex rule-definitions are needed to support the decision making process, i.e. to propose alternative areas with less pollution or close to the area of interest.

Class `EnvironmentalData` represents environmental measurements, i.e. observations from different sources regarding air pollutants, pollen, weather or any other measurable environmental aspect that is involved in the recommendation process.

Class `Request` structures the content of a user-request for DS. An instance of that type connects all these notions (user profile characteristics, the location of interest, the existing AQ measurements, and preferred activities) that are fed into the rule-based DSS for personalized recommendation provision.

Class `Recommendation` represents messages with fixed content, as defined by environmental experts for recommendation, together with details on: (a) which type of user such messages concern, and (b) under which AQ conditions they should be informed. An instance of that type may contain the actual message to be inferred to the user. Rules defined in the respective layer of the framework handle the matching between categories of users and defined recommendations.

3.2 Reasoning Over Domain Knowledge

For the semantic interpretation and inference of new knowledge, we implemented a rule-based reasoning layer that fully complies with the *SPARQL Inferencing Notation* (*SPIN*) framework, which is a well-established standard for representing SPARQL rules and constraints on Semantic Web models [22]. SPIN rules are linked to ontology classes and stored alongside the domain model as RDF triples, thus supporting a holistic, dynamic, semantically enriched approach which fits perfectly to the requirements of our defined DS problem.

In the proposed framework, the rule layer is implemented separately from the abstract schema (TBox) and the assertions (ABox), and the rules are grouped in two different levels. Thus, the reasoning process operates incrementally, where inferences from one step serve as input to rules of the following steps, moving from more general to more specific derivations. Such a distinction and hierarchy of rules facilitates the extensibility of the reasoning module, eliminating the intervention to high levels of rules when lower levels change, and vice versa.

In order to fully exploit SPIN's capabilities, we capitalize on the SPIN vocabulary (http://spinrdf.org/spl); among the available concepts, we adopt/extend three specific types: `spin:Functions`, `spin:Rule`, and `spin:MagicProperties`, which serve different functionalities (select/construct triples, reuse SPARQL blocks, etc.).

The first level of rules (109 in total) takes into account the schema and populated instances and produces low level derivations for the next level. These rules handle the transformation of age values to relevant age groups, of AQ observations to AQI scales and the classification of user profiles into those profile categories for which recommendations are available, on the basis of provided environmental experts' knowledge. Considering the rule-driven user classification process, a user is classified in one of the available basic profiles (subclasses of class `Person`), or in any permissible combination of them. Rules handle complex profiles automatically, by "downgrading" them into those combinations for which recommendations are defined. A user may be provided with more than one recommendation messages, with respect to its profile characteristics and the rule-based classification output.

The second level of rules (150 in total) takes into account preceding results from user- and AQI-related inferences, as input to relevant SPIN functions, which in turn generate the text of the most relevant recommendation per case from a list of experts' recommendations residing in the ontology. An example of the second level's SPIN rule is presented in Fig. 3, where a recommendation is formulated on the basis of preceding inferences: the user belongs to a specific combined category (triple #6) and the user's location was inferred to have a bad AQI (triple #11) at the time of request for DS. The user's preferred language (triples #8-9) is also considered for the final recommendation outcome, since multilingual messages are integrated in the ontology.

3.3 Communication with Ontology-Based Framework

We establish an interoperable communication between the decision support KB and external modules, with the implementation of RESTful services. Complex ontological definitions and rules are hidden behind the developed web services, which handle the

```
CONSTRUCT {
    ?person :isProvidedWithRecommendation ?personalized_rec .      #1
    ?personalized_rec a :LimitExposureRecommendation .             #2
    ?personalized_rec :hasDescription ?rec_description .           #3
    ?personalized_rec :hasDescriptionIdentifier "walking" .        #4
} WHERE {
    ?request :involvesPerson ?person .                             #5
    ?person a :Pregnant_SportsWalking_Person .                     #6
    ?person a :Person .                                            #7
    ?person :hasPreferredLanguage ?preferredLang .                 #8
    ?preferredLang :hasLanguageCode ?langCode .                    #9
    ?request :involvesLocation ?location .                         #10
    ?location :hasRelatedIndex :AQI_bad .                          #11
    BIND (BNODE() AS ?personalized_rec) .                          #12
    BIND ( :getPersonalizedRecommendation_langParam(?langCode, "bad",
:Pregnant_SportsWalking_Person) AS ?rec_description) . }           #13
```

Fig. 3. An example rule that combines inferred data and function's output to create new triples.

dynamic population of the ontology with actual data (AQ observations, user profile details, requests for DS) and the automated inference of personalized recommendations. Services were implemented in Java EE 7, with the adoption of Apache Jena framework (https://jena.apache.org/) for manipulating RDF graphs and the SPIN API (http://topbraid.org/spin/api/) for performing rule-based inference.

4 Test Cases and Inference Results

Our proposed ontology-based framework comprises the operational DS module of the hackAIR platform for generating meaningful QoL information, personalized according to the citizens' profile requirements. In this context, we demonstrate its functionality through the following indicative scenario: "*Valeria, a 32-year-old woman, pregnant with respiratory problems (asthma), queries the hackAIR application for information about existing AQ in an area where she usually goes for long walks. At the time of request, the PM_{10} values are extremely high (e.g. 150 μg/m^3)*". The process executed is given below:

(1) The request is sent from the hackAIR app to the ontology-based module through the supported web service;
(2) The user-profile details (Table 1-A) as well as the location of interest (Table 1-B) and AQ measurement (Table 1-C) are dynamically populated in the ABox according to the schema declared in the TBox;
(3) An instance of `Request` type is additionally formulated (Table 1-D) by integrates all involved information declared in the previous step;
(4) The recommendation module is triggered. SPIN rules are activated according to the level of suitability to the case. Results are produced dynamically (Table 1-E), moving from more general to more specific inferences;
(5) Inferences asserted as recommendations to the user, are dynamically extracted from the ontology and pushed back to the application for visualization.

Table 1. Instances in Turtle format (Details about the Turtle format are available at: https://www.w3.org/TR/turtle/) as populated/inferenced in our proposed framework

A. User profile

```
abox:Valeria
  rdf:type TBox:Person ;  tbox:hasAge 32^^xsd:integer ;
  tbox:hasGender tbox:female ; tbox:isSensitiveTo tbox:Asthma ;
  tbox:isPregnant "true"^^xsd:boolean ;
  tbox:hasLocation abox:location_V ;
  tbox:hasPreferredActivity tbox:walking_activity ; .
```

B. Location

```
abox:location_V
  tbox:hasEnvironmentalData abox:PM10EnvData_location_V .
```

C. Environmental data

```
abox:PM10EnvData_location_V
  rdf:type tbox:AirPollutantEnvironmentalData ;
  tbox:hasEnvironmentalDataType tbox:PM10 ;
  tbox:hasNumericalValue abox:PM10Value_location_V ; .
abox:PM10Value_location_V
  rdf:type tbox:AirPollutantValue ;
  tbox:hasUnit abox:microGramsPerCubicMeter ;
  tbox:hasValueValue "150.0"^^xsd:double ; .
```

D. Request

```
abox:request_XYZ
  rdf:type tbox:Request ;
  tbox:involvesEnvironmentalData abox:PM10EnvData_location_V ;
  tbox:involvesLocation abox:location_V ;
  tbox:involvesPerson abox:Valeria ;
```

E. Inferences

```
abox:Valeria
  rdf:type tbox:AdultPerson ;
  rdf:type tbox:PregnantFemalePerson ;
  rdf:type tbox:SensitiveHealthPerson ;
  rdf:type tbox:SportsWalkingPerson ;
  rdf:type tbox:Pregnant_Sensitive_Person ;
  rdf:type tbox:Pregnant_SportsWalking_Person ; .
abox:location_V tbox:hasRelatedIndex tbox:AQI_bad ; .
abox:Valeria
  tbox:isProvidedWithRecommendation [
      rdf:type tbox:LimitExposureRecommendation ;
      tbox:hasDescription "You should go for a walk in an area with
           cleaner air."@en ;
      tbox:hasDescriptionIdentifier "walking" ; ] ;
  tbox:isProvidedWithRecommendation [
      rdf:type tbox:LimitExposureRecommendation ;
      tbox:hasDescription "Consider avoiding any intense outdoor activ-
           ity in your area! The existing air quality might be harmful
           for your health."@en ;
      tbox:hasDescriptionIdentifier "general personalized" ; ] ; .
```

Focusing on the inference results presented in Table 1-E, we may distinguish the move from *general-to-specific* classifications, as those were produced by automatically executing different types/levels of implemented SPIN rules. First, the user is categorized as `PregnantFemalePerson`, `SensitiveHealthPerson` and `SportsWalkingPerson`, conforming to relevant information provided in the user profile (Table 1-A); the user is then categorized in two combined classes (`Pregnant_Sensitive_Person` and `Pregnant_SportsWalking_Person`) which have direct association to the available recommendations provided by the system. Considering these two specific subclasses that the user belongs to, and given the fact that the air quality index in the area of interest is inferred as `AQI_bad`, the system suggests *two* different recommendations, one related to the user's preferred activity (walking) and the other to the user's health sensitivity (pregnancy, asthma).

In a different scenario, the rule-based inference mechanism would follow the same reasoning sequence but would trigger different rules that correspond to the semantics behind the interpreted data. For example, different user profile characteristics or preferences would lead to different user profile classifications and, thus, different recommendation messages to the users, on the basis of existing AQ condition. Representative scenarios that were created within the hackAIR project, with support from the project partners and environmental experts, are demonstrated in [23].

5 Evaluation

For the evaluation of the proposed representation and reasoning framework, we focus on the following aspects: (i) the *consistency* of the provided results, by examining if the inferred recommendations comply with those targeted to be given through the classification and reasoning process; (ii) its *performance*, in terms of elapsed time when a request is submitted to the system. Unfortunately, a direct comparison of response times between the proposed framework and alternative approaches [5, 15] is not feasible, since there are no benchmarks to follow; systems have different complexity, demonstrate different functionalities, input or internal processes, target different recommendation outcomes, and the implementation details that are missing block the reproduction of identical experiments within our proposed context.

For the consistency checking task, the environmental experts and pilot users of the hackAIR project performed a thorough analysis of the reasoning process by examining the recommendation inferences for each ontology-represented use case. The evaluation showed a deviation from the planned system behavior in only two cases, due to wrong intermediate classifications. The problem was fixed by correcting the corresponding rules in the ontology.

Regarding the performance evaluation of the implemented DSS, we considered 11 problem (request) descriptions, differing in size and content: simple and complex profiles, with one or more involved users per request, with different profile characteristics, etc. For all use cases, the response time of the overall DS process ranged from 1.32 to 1.89 s (any differentiation in times depends mainly on the complexity of the rules performed per case), with an average time of 1.61 s. The evaluation ran on a computer with: Intel® Core™ i5-4690 K, x64-based processor at 3.50 GHz, with

16 GB RAM memory. Overall, our ontology-based DSS was proved to be a lightweight approach, complying with the needs for fast computations and response times of the reasoning service, avoiding any redundant system complexity.

6 Advantages of the Proposed Implementation

Through the extensive use of ontologies we achieve to create an integrated, stand-alone DS framework that is *self-identified* (structures the content and dynamically computes implicit values based on knowledge stated in the schema), and *self-triggered* (fires the reasoning process automatically and appropriate rules are executed when needed). The advantages of such an implementation are described below.

Uniformity: SPIN rules are defined as RDF triples alongside the ontology model. Such an approach prevents additional programming effort (rule-based inference is operated via the integrated SPIN engine) and facilitates the adoption of the DS module and external sources through implemented RESTful web services.

Flexibility: The multifaceted proposed framework permits multilingual definition of recommendations through the initialization of a single instance of `Recommendation` type per message. This can "carry" the same text in any of the supported languages, while an appropriate SPIN rule will return to the user the inferred message on his/her preferred spoken language. Different recommendation messages for the same use case are also easily supported, in a sense that the proposed framework can select either randomly, or by validating specific weighted factors, a personalized recommendation from a "bag of messages", all of which are defined as relevant for the examined case.

Modularity: The parameterization of implemented SPIN rules and functions allow complex schema and property declarations to be manageable in terms of maintenance, adoption and re-use; changing the body of a rule does not affect its use in other stages of the rule layer, while changing the parameters used/returned only alters its function signature. Moreover, SPIN enables the explicit definition of the rule execution order, giving priority to or isolating groups of rules under specified circumstances.

Extensibility: The hierarchical capabilities of the proposed framework facilitate the update and extension of rules and concepts supporting additional input from AQ, health or other related domain knowledge, without the need to change rules in other layers. More specifically, in our implementation, air pollutant measurements are converted via rules to corresponding AQI values; such data serve as input to subsequent levels of the reasoning process. If an additional air pollutant is to be added to the schema, rules will be enriched only for the task of converting numerical to nominal values, without any interference to AQI rules or SPIN rules interconnected to them.

Appropriateness: The combination of ontologies with SPIN rules, as an extra layer of rule-based reasoning, is perfectly suited for more advanced, rule-based inference. This sophisticated implementation provides the required expressiveness for the domain's knowledge representation and the respective inference mechanisms which can efficiently deal with the provision of user-profile driven recommendations.

7 Conclusions and Future Work

This paper presented a novel ontology-based framework, which integrates both the *static parts* and the *dynamic processes* of a DSS in a uniformly structured and semantically enriched decision support KB. The implementation is based on a 3-layered approach which consists of the following components: (i) the ontological schema of abstract concepts and relations in the domain of interest, (ii) the structured initialization of actual data in the ontology, and (iii) the rule-based mechanism for interpreting the semantics behind the stored data and generating the targeted recommendations to the users, with respect to their user profile characteristics. The developed schema and the mechanisms integrated in the proposed DSS framework, reflect the dynamics of the AQ domain from the citizens' perspective: in every request, the system combines the status of the environment and the specific characteristics of the requesting user, so as to produce simple personalized recommendations encompassing the direct influence of severe AQ conditions in the users' health and planned activities.

The proposed implementation highlights the strategic advantages of the use of ontologies on the basis of the multifaceted needs of DSSs: the layered structure and the semantic declaration of concepts and rules can efficiently handle the heterogeneity of represented data and facilitate the modularity and extensibility of the system. Moreover, the ontology-based implementation is domain-agnostic and can be exploited in different application contexts, following a relevant configuration mechanism of the targeted user profiles and recommendations.

As future work, we aim to further evaluate the satisfiability and coverage of the recommendation system from the users' perspective. We also aim to investigate the integration of fuzzy reasoning techniques for covering more complex relations (additional parameters, cumulative effects, etc.) Finally, we will propose an Ontology Design Pattern for the formal representation of the reasoning mechanism, and exploit new rule-based approaches with the adoption of SPIN's next generation called SHACL [24].

Acknowledgments. This work is partially funded by the European Commission under the contract number H2020-688363 hackAIR.

References

1. Karatzas, K., Kukkonen, J.: Quality of life information services towards a sustainable society for the atmospheric environment. In: COST Action ES0602 Workshop Proceedings (2009)
2. Zhu, L., Karatzas, K., Lee, J.: Urban environmental information perception and multimodal communication: the air quality example. In: Esposito, A., Hussain, A., Marinaro, M., Martone, R. (eds.) Multimodal Signals: Cognitive and Algorithmic Issues. LNCS (LNAI), vol. 5398, pp. 288–299. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00525-1_29
3. Michels, H., Mauè, P.: Semantics for notifying events in the affecting environment. EnviroInfo Conference, pp. 501–507 (2010)
4. Karatzas, K., Endregard, G., Fløisand, I.: Citizen-oriented environmental information services: usage and impact modelling. In: 19th EnviroInfo Conference, pp. 872–878 (2005)

5. Rospocher, M., Serafini, L.: An ontological framework for decision support. In: Takeda, H., Qu, Y., Mizoguchi, R., Kitamura, Y. (eds.) JIST 2012. LNCS, vol. 7774, pp. 239–254. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37996-3_16
6. Moumtzidou, A., Papadopoulos, S., Vrochidis, S., Kompatsiaris, I., Kourtidis, K., Hloupis, G., Stavrakas, I., Papachristopoulou, K., Keratidis, C.: Towards air quality estimation using collected multimodal environmental data. In: Satsiou, A., Panos, G., Praggidis, I., Vrochidis, S., Papadopoulos, S., Keratidis, C., Syropoulou, P., Liu, H.-Y. (eds.) IFIN/ISEM -2016. LNCS, vol. 10078, pp. 147–156. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50237-3_7
7. McIntosh, B.S., et al.: Environmental decision support systems (EDSS) development-challenges and best practices. Environ. Model. Softw. J. **26**(12), 1389–1402 (2011). https://doi.org/10.1016/j.envsoft.2011.09.009
8. Matthies, M., Giupponi, C., Ostendorf, B.: Environmental decision support systems: current issues, methods and tools. Environ. Model. Softw. J. **22**(2), 123–127 (2007). https://doi.org/10.1016/j.envsoft.2005.09.005
9. Raskin, R., Pan, M.: Semantic web for earth and environmental terminology (SWEET). In: Proceedings of the Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data (2003)
10. Buttigieg, P.L., Morrison, N., Smith, B., Mungall, C.J., Lewis, S.E.: The environment ontology: contextualising biological and biomedical entities. J. Biom. Semant. **4**(1), 43 (2013). https://doi.org/10.1186/2041-1480-4-43
11. Oprea, M.M.: AIR_POLLUTION_Onto: an ontology for air pollution analysis and control. In: Iliadis, L., Maglogiannis, I., Tsoumakas, G., Vlahavas, I., Bramer, M. (eds.) AIAI 2009. IFIPAICT, vol. 296, pp. 135–143. Springer, Boston (2016). https://doi.org/10.1007/978-1-4419-0221-4_17
12. Bicer, V., Tran, T., Abecker, A., Nedkov, R.: KOIOS: utilizing semantic search for easy-access and visualization of structured environmental data. In: Aroyo, L., Welty, C., Alani, H., Taylor, J., Bernstein, A., Kagal, L., Noy, N., Blomqvist, E. (eds.) ISWC 2011. LNCS, vol. 7032, pp. 1–16. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25093-4_1
13. Ruther, M., Bandholtz, T., Logean, A.: Linked environment data for the life sciences. In: 3rd International Workshop on Semantic Web Applications & Tools for the Life Sciences (2010)
14. Kontopoulos, E., Martinopoulos, G., Lazarou, D., Bassiliades, N.: An ontology-based decision support tool for optimising domestic solar hot water system selection. J. Cleaner Prod. **112**, 4636–4646 (2016)
15. Wetz, P., Trinh, T.D., Do, B.L., Anjomshoaa, A., Kiesling, E., Tjoa, A.M.: Towards an environmental information system for semantic stream data. In: 28th EnviroInfo Conference, pp. 637–644 (2014)
16. Wanner, L., et al.: Ontology-centered environmental information delivery for personalised decision support. Expert Syst. Appl. **42**, 5032–5046 (2015). https://doi.org/10.1016/j.eswa.2015.02.048
17. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquisition J. **5**(2), 199–220 (1993). https://doi.org/10.1006/knac.1993.1008
18. Turban, E., Aronson, J.E., Liang, T.P.: Decision Support Systems and Intelligent Systems, pp. 100–136. Pearson Prentice-Hall, Upper Saddle River (2005)
19. Suárez-Figueroa, M.C., Gómez-Pérez, A., Villazón-Terrazas, B.: How to write and use the ontology requirements specification document. In: Meersman, R., Dillon, T., Herrero, P. (eds.) OTM 2009. LNCS, vol. 5871, pp. 966–982. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-05151-7_16

20. W3C OWL Working Group: OWL 2 Web Ontology Language. W3C Recommendation (2012). https://www.w3.org/TR/owl2-overview/
21. Falco, R., Gangemi, A., Peroni, S., Shotton, D., Vitali, F.: Modelling OWL ontologies with graffoo. In: Presutti, V., Blomqvist, E., Troncy, R., Sack, H., Papadakis, I., Tordai, A. (eds.) ESWC 2014. LNCS, vol. 8798, pp. 320–325. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11955-7_42
22. Knublauch, H., Hendler, J.A., Idehen, K.: SPIN - overview and motivation. W3C Member Submission (2011). https://www.w3.org/Submission/spin-overview/
23. hackAIR Consortium. D4.2: Semantic integration and reasoning of environmental data (2017). http://www.hackair.eu/wp-content/uploads/2016/03/d4.2-semantic_integration_and_reasoning_of_environmental_data.pdf
24. Knublauch, H., Kontokostas, D.: Shapes Constraint Language (SHACL). W3C Recommendation (2017). https://www.w3.org/TR/shacl/

Decision Support Systems Serving the Public

Critical Events and Critical Infrastructures: A System Dynamics Approach

Stefano Armenia[1], Georgios Tsaples[2(✉)], and Camillo Carlini[1]

[1] Department of Computer, Control and Management Engineering, Sapienza University of Rome, via Ariosto 25, 00185 Rome, Italy
{armenia,carlini}@dis.uniroma1.it

[2] Department of Business Administration, University of Macedonia, Thessaloniki, Greece
gtsaples@gmail.com

Abstract. Critical events like natural disasters or terrorist attacks have evolved to a realistic concern over the last decade. Anticipation and timely decision-making are fundamental to ensure positive outcomes, stability and safety especially with concern to Critical Infrastructure. The purpose of this paper is to promote a new framework for efficient and effective crisis management in the area of Critical Infrastructures. The main effort of the project has been to use simulation models with the purpose of investigating, reproducing and representing the interdependencies among CIs while stressed by critical events. The results of the simulation demonstrated that preparation in cases like a flooding, can prove extremely useful and make an extreme situation more manageable. Finally, elements of the human behaviour, which are the most uncertain and difficult to simulate, are the most important and can make the difference between a disaster and a manageable situation.

Keywords: System Dynamics · Simulation · Interdependencies
Critical Infrastructures · Critical events

1 Introduction

In recent years the debate on awareness about the fundamental role played by Critical Infrastructures (CI in the rest of the text) in today's society has spread from the limited circles of professional insiders to a high decision-making level, bringing the issue in the front lines of national and transnational agendas.

In a rapidly changing world, critical events (either natural like earthquakes and flooding or man-caused like technology failures and terrorist attacks) have evolved from a rare and unpredictable occasion to a concrete and realistic concern for stakeholders at both private and public levels [1]. Three elements are recurrent during critical events: first there is the involvement of a complex infrastructure system, which facilitates the spread of negative effects to other parts of the community or the country, second is the surprise factor, which is determined by the type of the critical event and can take various degrees of manifestation and finally the short decision time available to react and manage the evolving situation.

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 55–66, 2018.
https://doi.org/10.1007/978-3-319-90315-6_5

In this context, anticipation and timely decision-making are fundamental to ensure positive outcomes, stability and safety. Properly evaluating future events, promptly detecting emergencies and timely assessing the impact of different policy choices before they are implemented are necessary for a successful management of a crisis [2].

A crisis is defined as an event with characteristics of:

– High uncertainty and unfamiliarity
– Low probability of occurrence
– Requirements of a rapid response
– Serious threats to survival
– The need for a decision that will either result in a positive or a negative change [3].

As a result, crisis management is concerned with dealing with effects of potential threats before, during and after the event. However, the success of a crisis management is hindered by the interconnectedness of CIs. A critical event can generate domino effects that can cripple entire regions, which explains the latest shift in the importance of CIs to the national and transnational agendas.

The purpose of this paper is to present the CRISADMIN project (***Cr**itical **I**nfrastructure **S**imulation of **Ad**vanced **M**odels on **I**nterconnected **N**etworks resilience*) cofounded by the Prevention, Preparedness and Consequence Management of Terrorism and other Security related Risks Programme of the EU-DG Home Affairs (HOME/2011/CIPS/AG/4000002116), has aimed to promote a new framework for efficient and effective crisis management in the area of Critical Infrastructures. The main effort of the project has been to use simulation models with the purpose of investigating, reproducing and representing the interdependencies among CIs while stressed by critical events. The developed prototype has been designed to allow decision-makers to analyze the interdependencies among CIs, the modalities through which they get affected by unpredictable, catastrophic events as well as to investigate the impacts of potential intervention countermeasures or prevention policies.

The CRISADMIN tool can be used in contexts where standard analysis is rendered difficult by the wide range of available data and/or relations in place. In particular, it could be especially useful in situations where aspects of human behavior influence not only the operation of CIs, but also how negative effects can propagate in case of a critical event.

The rest of the paper is structured as follows: Sect. 2 is focused on the methodology that was used for the CRISADMIN, while in Sect. 3 the structure of the developed model is presented. In Sect. 4 there is an extensive description of various scenarios, their results and what are the implications of the designed (simulated) policies. Finally, conclusions and a discussion are presented in Sect. 5.

2 Methodology

This section is focused on the methodology that was used to develop the simulation model and the details and justifications that follow the specific choice.

Critical Infrastructures are inherently complex. System complexity can be distinguished by several factors: the number of modes of the system, the number of relations among the various sub-systems [4]. Moreover, the decision-making process is dynamic: it requires more than one decisions, decisions are interdependent and the environment changes in the course of time, either as a result of the decision or independently [5, 6]. In this context, decision-makers may have good decision rules, but they might fail to apply them consistently [7].

To help decision-makers to understand the errors in their judgments and make more effective decisions, computer-simulation models are used. They are adequate representations of reality, which offer to decision-makers the opportunity to gain experience and test world-like responses [6]. Using simulation offers a variety of advantages:

- Simulation allows the analysis of large scale systems
- It is easy to use
- It allows the experimentation with different scenarios and "what if" analyses
- It can account for different perspectives [8].

System Dynamics [9, 10] is a computer-based methodology that makes use of formal models to enable enhanced understanding of complex systems over time. This understanding provides insights to decision-makers on how a policy might change (favorably or not) the state of the system under study [11].

System Dynamics is capable of inherently taking into account randomness and interdependency, which both characterize the behaviour of real-life business environments. Furthermore, it allows for the inclusion of those "soft" variables typical of the interrelated social systems, which generally are not taken into account in most of the linear modelling techniques still used nowadays. The idea behind the System Dynamics-approach is that, if "a system structure defines its behaviour" [10], then by being accurate in analysing and determining the interrelationships among various parts of the system, it could be possible to define accurately the structure of the problem under study and this would ultimately bring an increased understanding of the dynamics of the system. Thus, the SD replicates dynamic business reality with the power to "look into the future" and to understand the impact on multiple key metrics. Additionally, simulation allows the user to capture the specific variability of multiple processes and ultimately provides results, which are orders of magnitude more accurate than deterministic analysis [12].

Finally, the elaborated System Dynamics model can be embedded in an Interactive Learning Environment, which will constitute the end-users' interface for testing the impacts of critical events on CIs.

3 Model Structure

The general structure of the System Dynamics model- at a high level- is depicted in the Causal Loop Diagram of Fig. 1.

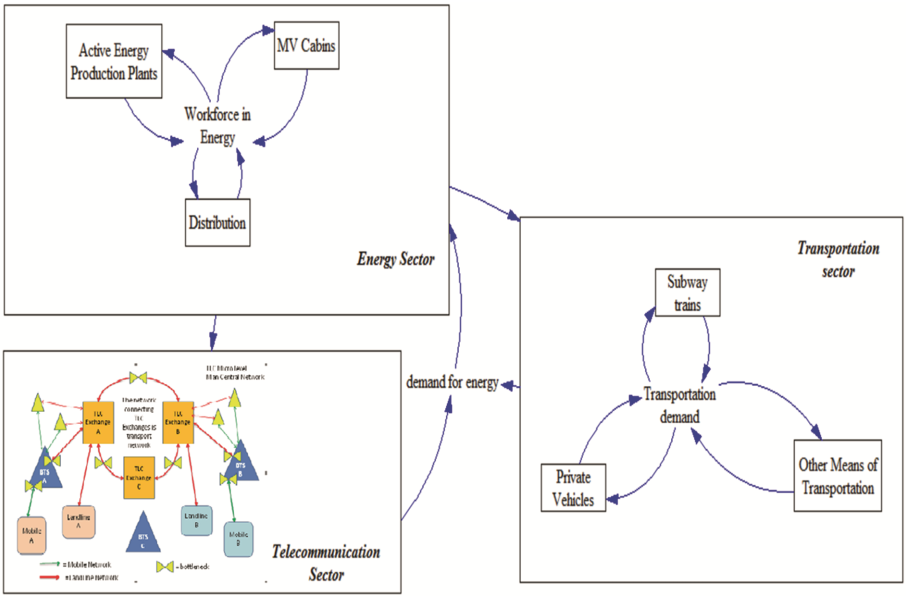


Fig. 1. General Causal Loop Diagram

Energy production facilities produce High Voltage (HV), which is transferred to the distribution area. The connection between them is the Medium Voltage (MV) Transformation cabins. Consequently, the MV energy is transformed into Low Voltage and distributed to the metropolitan area. Each of the structures of the energy sector can malfunction either normally or due to a critical event and can be repaired with the appropriate allocation of the energy-sector-workforce. The energy sector determines the functionality of the rest of the Critical Infrastructures of the CRISADMIN model. More specifically, in the Telecommunication sector (TLC), the Base Transceiver Stations (BTS) that handle the mobile demand, are directly affected by the energy sector. Similarly, the subway network is connected to the Medium-Voltage part of the energy sector. A malfunction in the energy sector will affect the performance/functionality of the transportation sector, since the subway will not function properly. Thus, the passengers of the metropolitan area will seek new means of transportation. However, new means of transportation means new travel patterns (and increased traffic flow) that will affect the energy workforce ability to reach the points of damage (in the energy sector) and start the repairs. Both the energy and the transportation sector affect the health sector of the model. For a more detailed description of the CRISADMIN qualitative model [13, 14].

The underlying System Dynamics model captures an urban environment (the values of which can be changed to create different urban environments) with four CIs: Energy, Telecommunications, Transportation and Apparatus. The latter two are the sectors that are directly influenced by the uncertain human behavior, thus a more detailed description of their main elements will be provided in the next paragraphs.

3.1 Transportation

The purpose of the Transportation sector is to describe the iterations of the main infrastructure choices represented by Subway and Other Means of Transportation (OMT) for the public transport and by private transport (PT). In order to better explain these relationship, it is useful to explain the choice mechanism implemented that allows people to choose which type of means of transport to use, in accordance with the relevant transport service level.

The reference population area must decide whether to travel by subway, by bus (OMT) or by car (PT). This choice, in the model, depends on a dynamic variable factor represented by the corresponding service level (with values ranging between 0 and 1) of the 3 sub-models contained within the transport infrastructure sector [15].

The subway behavior depends on the number of lines, stations and trains that are active (and not inactive due to normal malfunctions and/or damages due to critical events) and the number of people who use the subway system. For the calculation of the Subway Service Level (SL) is defined as the number of people who arrive at the destination divided by the number of people who asked for a train.

The other sectors of the transportation system have similar elements as the subway sector. To avoid repetition, we mention only one variable that is important not only for the transportation sector but the entire model. The variable is Key Performance Indicator of the traffic. To calculate the variable, the relationship between the Total Vehicles and the "Available Equivalent n° of street KMs in Ref Area" is taken into account. As mentioned before, the "Traffic KPI" is a very important variable, whose changes affect the entire CRISADMIN model.

Finally, the workforce responsible for the repairs on the damaged materials of the sector, moves dynamically where is mostly needed.

3.2 Apparatus

The Apparatus section contains the elements of the model that are the most difficult to simulate: Health services and human behavior.

Health Management

The Health sector is represented as a supply chain in which patients are entered (after injury), are treated in the various stages of the chain depending on their injuries and leave the chain. Big part of the Health sub-model is the management of the ambulances and the medical personnel. Both move dynamically to where they are mostly needed and their successful operation determines the pressure under which the health management functions.

Human Behavior

Acritical event induces behavioral changes in the people that are part of the system being in crisis as well as in those that are supposed to manage and solve such crisis. In making selections, the main focus was the human behavior in social systems in the response phase [16, 17]. The elements of human behavior are separated into to main sub-models: the behavior of the authorities and of the citizens. These aspects of a system are represented by variables that are called "soft". Soft variables relate to attributes of human behavior or effects that variations in such behavior produce. Numerical data are often unavailable or non-existent for soft variables. However, such variables are known to be critical to decision making and, therefore should be incorporated into system dynamics models.

Authorities

This section describes how information flows from the place in which the critical event has been detected to the superior hierarchical levels of command. Information is collected from local monitoring activities. To be effective, the localif the water level does not fall below the alarmly collected information needs to be transformed into a locally authorized and organized reaction.

3.3 Critical Events

In the context of the CRISADMIN ILE different types of flooding can be accounted for (along with alternate critical events [18]). The critical event of flooding can be defined by every user independently to study different scenarios. The main variables that determine the scenarios are: the millimeters of rain per day, the water level considered as flooding, the capacity of the river banks and the weight of the monitoring that determines the prevention that increases the possibilities to intervene.

3.4 Policies and Countermeasures

Several countermeasures have been designed and developed for the simulation model. The policies are nothing more than variables in the model that change the value/behavior of basic variables. For CRISADMIN, five groups of variables that can influence the model output and be used as policies/interventions were identified and tested.

- **Policies in the energy sector:** These policies are focused on re-allocating the energy appropriately to the other sectors and prioritize those that are in greater need for energy.
- **Policies in the transportation sector:** These policies are designed to help the transportation sector operate as smoothly as possible after a critical event.
- **Policies for the availability of resources:** The rationale behind those interventions is that in cases of emergencies, more resources (either human or material) are usually necessary.
- **Policies on people's behavior:** The rationale behind those policies is to try and control people's behavior after a critical event. For example, if people start behaving "normally" sooner, then chaos might be contained, thus ensuring that less people are injured and the pressure in the health organization remains manageable. However, these policies are extremely difficult to materialize (how can someone control effectively people's behavior after a critical event?) and extremely uncertain (how can a "real" value be assigned to the intervention?).

- **Policies on the authorities' behavior:** These policies/interventions are focused on how the authorities can be organized faster to avoid the most severe/crippling effects of a critical event. Similar to the previous type of policies, they are difficult to materialize and uncertain. However, they can be very effective, especially if they are applied before a critical event.

3.5 Interactive Learning Environment

To facilitate decision-makers and engage them in an active learning process, an Interactive Learning Environment was designed and developed for the CRISADMIN model. Figure 2 demonstrates the main dashboard from which users can interact, simulate different scenarios and/or critical events and test counter-measures in a consequence-free environment.

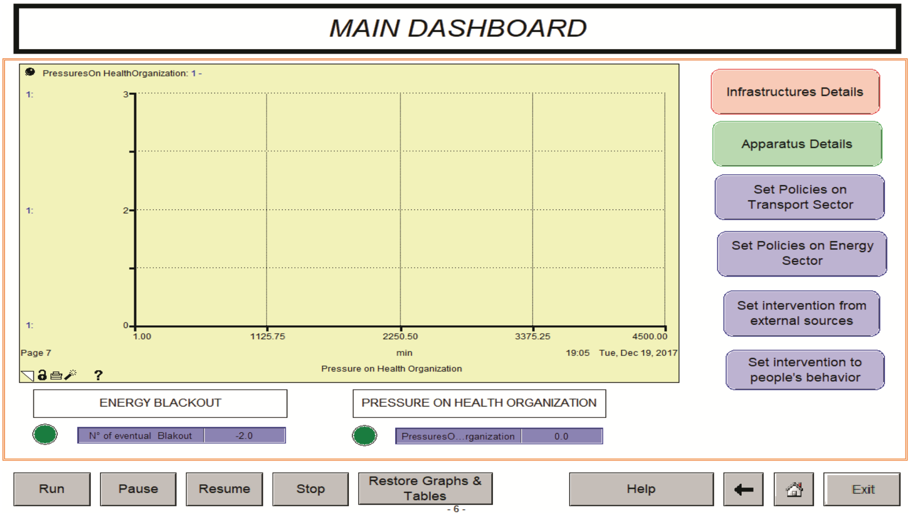


Fig. 2. The main dashboard of the interactive learning environment

4 Scenarios, Simulations and Results

The previous sector dealt with the explanation of the structure of the CRISADMIN model. The detailed explanation served a specific cause: to understand the results, the path that lead to them is necessary. Each modelling process consists of several steps and to fully appreciate the insight a System Dynamics model has to offer, one should be able to understand the theoretical background, the reasoning behind the qualitative model, the structure of the quantitative model and finally, the rationale behind this specific structure.

4.1 Flooding Event

The simulated scenario deals with an excessive amount of rain that gradually declines. For the purpose of the simulation, it was assumed that in the scenario "water bomb", the shape of the function that determines the amount of rain follows that of Fig. 3. What the graphical function simulates, is a constant, heavy rain for one day, after which it starts to slow down and stops completely half way in the second day.

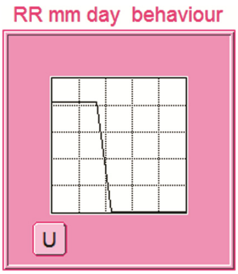


Fig. 3. Shape of the function that determines the rain in the "water bomb" scenario

Furthermore, it was assumed that the main points of impact would be the subway stations and the Low-Voltage cabins. The rationale behind the choice was to examine how the interdependencies of the various sectors would affect their operations/functions if a critical event would be applied to two major sectors.

In the "water bomb" scenario, it was assumed that heavy rain falls in the simulated metropolitan area, which gradually declines. As points of impact, the subway stations and the low-voltage cabins were defined.

The structure of the model dictates the expectations for the model outputs. The flooded subway stops people from using the metro system, leading them to look for other means of transportation. However, the larger the part of the population that will use private transportation-in addition to the flooded streets- is expected to lead to traffic congestion.

Moreover, the problems in the low-voltage cabins are expected to create problems in the distribution of electricity, thus leading to energy Blackouts. In addition, the critical event is expected to have an effect on people's behaviour. Since the change in people's behaviour is represented as an increased demand for telephone calls and in addition with the problems that are expected in the telecommunication sector due to the energy shortage, a congestion on the telecommunication lines are possible.

Finally, the critical effect has direct effects on people. The number of injured people will increase. Adding the problems from the energy shortage, we expect to see an increased pressure in the health-organization sector of the model.

The increased water level and the incapability of the river banks to hold that much water, causes a flooding. The amount of water will start decreasing after the set scenario period (one and a half days). As a result, the number of active low-voltage energy cabins is reduced, leading to blackouts (Fig. 4).

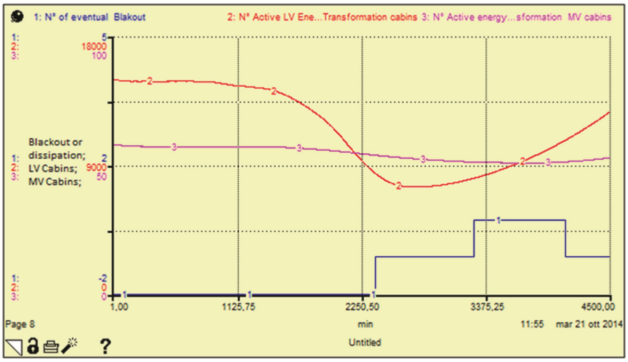


Fig. 4. Effect of “water bomb” on the energy sector

The shortage of energy, in combination with the flooding of subway stations, causes the de-activation of subway lines and the disruption of the function of the subway (Figure). Furthermore, the repair of the lines cannot start until the water level has fallen below the alarm threshold and this only occurs at the final stages of the simulation time (circle number two on Fig. 5).

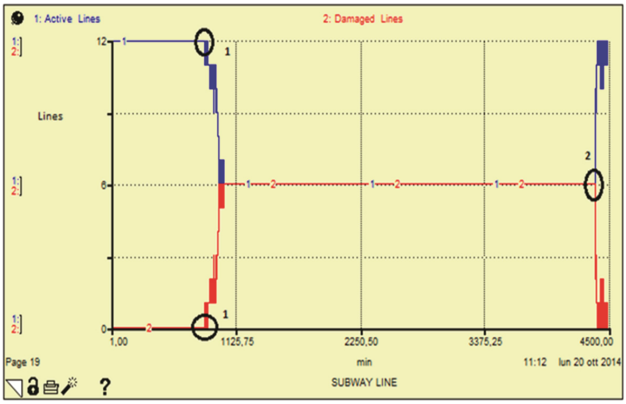


Fig. 5. Effect of the “water bomb” on the subway

As a result, people start looking for different means of transportation and their only output is private transport or other means of transportation (buses). Consequently, the increased number of vehicles on the (flooded streets) leads to a traffic congestion and an increased traffic KPI (Fig. 6 on the left). The increased traffic KPI in addition to the problems due to energy shortage leads to an increase pressure for the health organization (Fig. 6 on the right), since the injured people either cannot reach the hospital (high traffic KPI) or cannot be treated at the hospital (shortage of the necessary energy). Interestingly, the pressure shows signs of increase at the end of the simulation time. This can be explained by the nature of the flooding. It has cumulative effects, since nothing can be

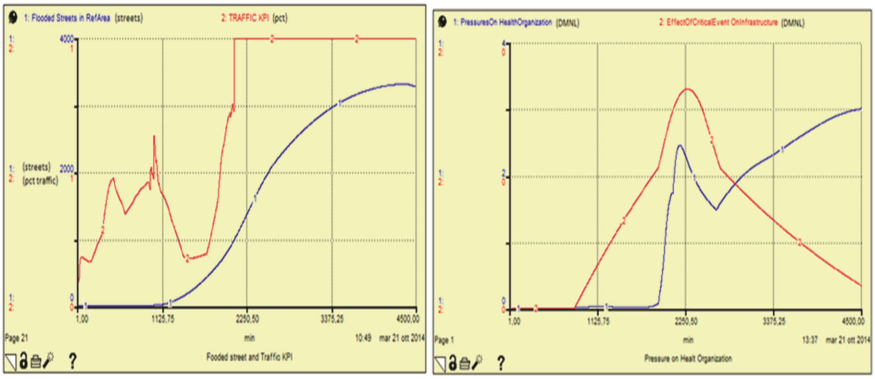


Fig. 6. Traffic KPI (left) and pressure on the health organization (right)

restored if the water level does not fall below the alarm threshold (and that occurs at the end of the simulation time).

Finally, the effect of the critical event on people's behaviour is manifested through an increased demand for telephone calls. The increased demand in addition to the energy shortage, leads to a congestion in the telecommunication lines (Fig. 7).

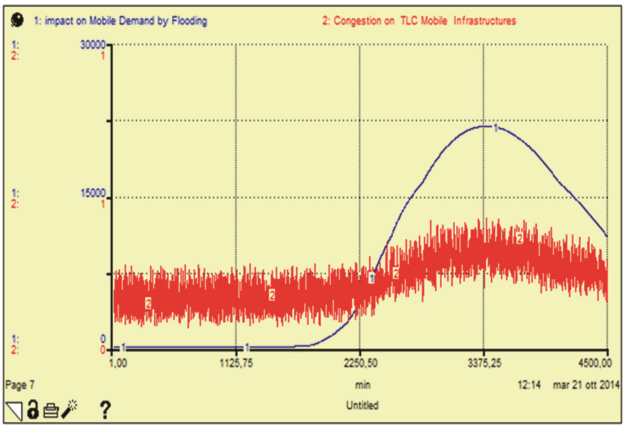


Fig. 7. Impact of the "water bomb" on the telecommunications

4.2 Indicative Policies and Countermeasures

The scenarios offered insights into the dynamics and degrees of connectedness of the various infrastructure sectors. However, the System Dynamics methodology can also be helpful in the area of policy design and testing.

For example, policies on the authorities' behavior are focused on how the authorities can be organized faster to avoid the most severe/crippling effects of a critical event. Similar to other types of policies they are difficult to materialize and uncertain. However, they can be very effective, especially if they are applied before a critical event. For example, if it can be made sure that the authorities perceive the danger of a flooding sooner–and take the appropriate actions, then the water level is reduced below the alarm threshold faster.

5 Conclusions

Decision makers need to understand the consequences of policy and investment options before they enact solutions, particularly due to the highly complex alternatives available for protecting the nation's CIs in today's threat environment. An effective way to examine trade-offs between the benefits of risk reduction and the costs of protective action is to utilize an Interactive Learning Environment that incorporates threat information and disruption consequences in quantitative analyses through advanced modelling and simulation. The proposed System Dynamic approach can provide decision makers with a useful means to understand and evaluate some of the expectable risks triggering CIs.

The CRISADMIN approach can be easily used in contexts where standard analysis is made difficult by the wide range of available data and/or relationships in place. In particular, it would be specifically helpful in those systems highly influenced by "soft" variables, connected to human behaviour.

The systemic approach, which closely follows the Systems Thinking & System Dynamics Methodology prescriptions, has allowed for a simple yet very effective representation of such context, with the identification of those parameters that, in a "domino effect", influence the behaviour of the whole interconnected system. The sub-systems described in the previous paragraphs, relevant to the sample context, constitute the second phase of the CRISADMIN Project, with the entire context being integrated right in these days into a quantitative model, with interdependencies between CIs being explicitly modelled and with aspects belonging to each Data Domain (including the "social" cross domain) being reasonably accounted for. According to the type of event selected, the most concerned variable of each subsystem will be appropriately stressed and the model will produce the various scenarios to be analysed.

Thanks to the work performed along with the CRISADMIN project development, it has been possible to identify the main interdependencies among systems' parameters, the main interconnections among critical infrastructures and the main variables influencing the evolution of critical events.

The results of the simulation demonstrated those characteristics that facilitate a crisis management to be successful. Mainly, preparation in cases like a flooding, can prove extremely useful and make an extreme situation more manageable. Secondly, the correct and fast allocation of the correct type of personnel is crucial to reduce the number of fatalities and restore the situation back to normal as soon as possible. Finally, elements of the human behaviour, which are the most uncertain and difficult to simulate, are the most important and can make the difference between a disaster and a manageable situation.

References

1. Hinssen, P.: The New Normal. Mach Media NV, Gent (2010)
2. Mureddu, F., Osimo, D., Misuraca, G., Onori, R., Armenia, S.: A living roadmap for policymaking 2.0. In: Sonntagbauer, P. (ed.) Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling, p. 433. IGI Global (2014)
3. Sayegh, L., Anthony, W., Perrewe, P.: Managerial decision-making under crisis: the role of emotion in an intuitive decision process. Hum. Resour. Manag. Rev. **14**(2), 179–199 (2004)
4. Gonzalez, C., Vanyukov, P., Martin, M.: The use of microworld to study dynamic decision making. Comput. Hum. Behav. **21**(2), 273–286 (2005)
5. Edwards, W.: Dynamic decision theory and probabilistic information processings. Hum. Factors: J. Hum. Factors Ergon. Soc. **4**(2), 59–74 (1962)
6. Quadrat-Ullah, H., Karakul, M.: Decision making in interactive learning environments: towards an integrated model. J. Decis. Syst. **16**(1), 79–99 (2007)
7. Hsiao, N., Richardson, G.: In search of theories of dynamic decision making: a literature review. In: 17th International Conference of the System Dynamics Society, Wellington, New Zealand, p. 2012 (1999)
8. Zhang, P., Peeta, S.: A generalized modeling framework to analyze interdependencies among infrastructure systems. Transp. Res. Part B: Methodol. **45**(3), 553–579 (2011)
9. Forrester, J.W.: Industrial dynamics. J. Oper. Res. Soc. **48**(10), 1037–1041 (1997)
10. Sterman, J.: Business Dynamics: Systems Thinking and Modeling for A Complex World, vol. 19. Irwin/McGraw-Hill, Boston (2000)
11. Santos, S., Belton, V., Howick, S.: Adding value to performance measurement by using system dynamics and multicriteria analysis. Int. J. Oper. Prod. Manag. **22**(11), 1246–1272 (2002)
12. Tsaples, G., Armenia, S.: Studying pension systems and retirement age: a simple system dynamics model for a complex issue. Int. J. Appl. Syst. Stud. **6**(3), 258–270 (2016)
13. Armenia, S., Carlini, C., Saullo, A., Assogna, P.: From theory to practice: taking up a theoretical framework on critical infrastructures modeling through a system dynamics approach. In: Proceedings of the 14th Conference of the European Academy of Management (EURAM) (2014)
14. Cavallini, S., d'Alessandro, C., Volpe, M., Armenia, S., Carlini, C., Brein, E., Assogna, P.: A system dynamics framework for modeling critical infrastructure resilience. In: Butts, J., Shenoi, S. (eds.) ICCIP 2014. IAICT, vol. 441, pp. 141–154. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45355-1_10
15. Radzicki, M., Sterman, J.: Evolutionary economics and system dynamics. In: England, R. (ed.) Evolutionary Concepts in Contemporary Economics, pp. 61–89. University of Michigan Press, Ann Arbor (1994)
16. Quarantelli, E.: Disaster related social behavior: summary of 50 years of research findings. Disaster Research Center (1999)
17. Tierney, K.: The social and community contexts of disaster. J. Nerv. Ment. Dis. **179**(8), 514–518 (1989)
18. Armenia, S., Tsaples, G., Carlini, C.: Interactive learning environments for crisis management through a system dynamics approach. In: Proceedings of the 15th European Academy of Management (EURAM) Conference (2015)

How to Model Stakeholder Participation for Flood Management

Marta Dell'Ovo[1(✉)], Francesca Torrieri[2], and Alessandra Oppio[3]

[1] Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, via Bonardi 9, 20133 Milan, Italy
marta.dellovo@polimi.it

[2] Department of Industrial Engineering, Università di Napoli Federico II, Corso Umberto I, 40, 80138 Naples, Italy
francesca.torrieri@unina.it

[3] Department of Architecture and Urban Studies, Politecnico di Milano, via Bonardi 3, 20133 Milan, Italy
alessandra.oppio@polimi.it

Abstract. Stakeholders participation for Flood Risk Management is a key factor for the improvement of policy and decision's quality of and to create consensus. Nowadays there are many studies on this topic aimed to take into consideration the involvement of stakeholders in different phases of the process and with the use of different procedures. In Italy the situation seems to be critical compared to the international panorama, since there are no regulation or protocols to prevent disaster or repair the damage. The paper proposes a critical overview of methodologies able to engage stakeholders in decision-making process with a detail on case studies focused on the Flood Risk Management. Different aspects will be investigated and compared in order to outline considerations and possible conclusions.

Keywords: Flood risk management · Stakeholder analysis
Geographic information system · Environment assessment

1 Introduction

The European Union underlines the importance of public and stakeholders participation in flood risk management (EU 2014) due to the complexity and uncertainty of the decision making process and the multiplicity of subjects involved. Actually, participation is considered a key instrument for involving public administrations, interest groups and individuals in the development of plans in order (i) to improve the quality of policies and decisions, (ii) to protect legitimate interests and (iii) to create the conditions for a large consensus among stakeholders on territorial development. Moreover, the stakeholders' involvement from the early stages of the planning process helps to reduce conflicts on the alternative solutions proposed for mitigate the risk and to integrate the "expert" knowledge with the "non-expert" knowledge, based on values and needs of the local populations. The participation of civil society increases the awareness about the risk [1] and induces a change of the behavioral approach to risk [2].

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 67–75, 2018.
https://doi.org/10.1007/978-3-319-90315-6_6

The idea of "public participation" has been introduced by the Water Flood Directive (2000/60/EC) and recalled by the Flood Directive 2007/60, where three kind of participation have been mentioned: (i) Information supply; (ii) Consultation; and (iii) Active involvement or participation in the decision making process. The Floods Directive (Article 10) states that "Member States shall encourage active involvement of interested parties in the production, review and updating of the flood risk management plans".

According to the CIS Guidance Document N.8, public participation is developed at three different levels: providing information, implementing consultation and encouraging active involvement for defining risk prevention and mitigation alternatives. A multilevel approach is recommended, including national, regional and local government organizations, in order to shift between vertical and horizontal approaches for modeling decisions, choice and evaluating. Not only institutions and stakeholders in the field of water management must be involved, but also institutions and stakeholders in the field of spatial planning, crisis management (civil protection) and/or economic sectors.

Given these premises, the paper proposes a review of the principal studies developed on stakeholder analysis/engagement and flood risk management in order to define a social integrated evaluation model for flood damages.

To this aim the first part of the paper presents a review of the literature on modeling the stakeholders' participation for flood management from different points of view, by analyzing experiences done in different countries with respect to the types of stakeholders, the techniques used for their involvement as well as the level of participation/ interaction. In the second part a social integrated flood damages valuation framework is described. Finally, the conclusions point out some operational recommendations.

2 Modeling the Stakeholders' Participation for Flood Management: A Literature Review

2.1 Overview the Stakeholders' Participation

According to Freeman [3], stakeholders are individuals or groups who affect, or are affected by an organization and its activities. Since that time the issue of stakeholders' engagement has grown emblematically according to the instance of deepening democracy within a sustainability perspective. Most of the research on stakeholders has been developed in four sub-fields: normative theories of business; corporate governance and organizational theory; corporate social responsibility and performance and strategic management [4].

Harrison and St. John [5] have given a substantial contribution to the latter, by focusing on the stakeholder approach as a strategic tool for analyzing, prioritizing and envisaging directions based on values. In order to use strategically the stakeholders' engagement, the first step is to provide a map. There is no generic list of stakeholders for all the decision problems.

Stakeholders should be aware of their role and their importance in achieving a result or choosing a scenario among a set of alternatives since "the outcome of a decisional process depends on the actors" [6].

Dealing with complex systems such as the Flood Risk Management, the choice of the right categories of stakeholder becomes fundamental and the methodology to use to assess their opinion and to stimulate an active interplay becomes even more important in order to reach a general consensus and to develop the most successful strategy.

Many fields of research and operative research are investigating the issue of the stakeholders' participation and how to involve them in decision-making processes. In particular, the adoption of participatory paradigm become fundamental in environmental decision problem where specific competences and knowledges are required [7].

There are several methodologies able to analyze actors involved, their connections, the level of power and interest and also how to interact with them. Brown et al. in 2011 [8] in order to solve a problem of natural resource management, consulted stakeholders to assess their preferences on a set of criteria previously defined with the aim of comparing different scenarios. Once identified groups of stakeholder, the work proceeded in a three-stage process. During the first phase groups of stakeholder have been met separately, the second one consisted in presenting to the other groups different preferences resulted from the meetings and in the third one groups have been brought together in workshops to create a general consensus. Here the participation was reduced to the last phase of the process, for the criteria weight elicitation, while stakeholders could be, in other cases, necessary to structure and identify possible strategies. Rockloff and Lockie [9], in fact, for the coastal management, interviewed face-to-face stakeholders in order to explore their expectations, then this information have been used to frame social maps aimed to identify strategies and to draw network among them. In 2010, Nordström et al. [10], combined the Multicriteria Decision Analysis (MCDA) and stakeholder analysis to solve participatory forest planning. Actors have been engaged in the whole process, identifying, first of all, the relevant stakeholders and the extent of their participation, then structuring the set of criteria according to their objectives and generating alternative solutions considering their preferences and finally, the elicitation has been performed by the pairwise comparisons procedure of the AHP in order to obtain a rank of alternatives. Considering the paper described and the literature about participatory processes, it is clear the importance of identifying the correct group of stakeholders, avoiding to exclude some categories. Lim and Bentley [11] proposed a novel method that uses a five types of model social network able to represent different types of stakeholder activity with the aim of understanding their involvement and choose the most suitable to be sought. Also Aburto et al. [12] for the governance of marine resources, investigated the stakeholder participation focusing their research on understanding and identifying people who should be involved in the sea council using two participatory workshop.

Therefore, it is possible to summarize that for decision-making process where multi-stakeholders are involved or should be taken into consideration, actors have to be defined carefully and also their level of importance and, according to the field of interest, the most suitable methodology of interaction has to be applied.

2.2 The Role of the Stakeholders in the Field of Flood Management

The field investigated by the paper concerns the flood management, and, in particular, the engagement of stakeholders in these kind of processes. The literature review is going to be focused on specific issues related to this topic aimed to categorize the level of participation and the level of interaction of stakeholders in the case studies under investigations.

The literature review has been carried out through the electronic database Scopus. The articles have been selected based on a keyword research in which the following combination have been used: "stakeholder participation" and "model" and "flood".

A total of 17 papers have been found and then a further sorting has been provided according to the abstracts. 15 papers have been considered relevant and suitable for the aim of the research. For each articles selected the following information has been recorded (Table 1):

- Author(s);
- Year;
- Decision problem;
- Case study;
- Category of stakeholder;
- Type of analysis;
- Use of Spatial analysis.

Table 1. Literature review on stakeholders' participation for flood Management

Author	Year	Decision problem	Case study	Category of stakeholders	Type of analysis	Use of spatial analysis
[7]	2016	Participatory approach to identify intervention options	UK	Experts and local stakeholders	Bayesian networks - workshop	No
[13]	2016	Integration of transportation and environmental planning	California (USA)	Agencies and common people	Web-based survey	No
[14]	2015	Examination of the causes and effects of typhoon disasters	China	Government and common people	Multi-stakeholder participation and vertical– horizontal coordination	No
[15]	2015	Web-based support tools for collaboration in FRM	UK - Germany	General public, planners and government organization, emergency managers, and flood management professionals	Web-based decision-making environment. Workshop	Yes
[16]	2014	Determining flood investment strategies	UK	Government and expert	Agent-based modelling (ABM)	No
[17]	2014	New model techniques to support decision-makers in urban water management	The Netherlands	Policy makers from the municipality, regional water board, province and fire department	Interviews and Two work sessions	Yes
[18]	2013	Innovative web-based virtual environment for flood information sharing and dissemination	Romania	Citizens, decision makers, authorities, specialist, modelers	Citizens/stakeholders feedback on the web platform	Yes

(*continued*)

Table 1. (*continued*)

Author	Year	Decision problem	Case study	Category of stakeholders	Type of analysis	Use of spatial analysis
[19]	2013	GIS framework (Polyscape) designed to support landscape management	UK	Local stakeholders (farmers, environmental managers and policy makes)	Interactive use of the tool	Yes
[20]	2013	Design of the Hungarian flood insurance system	Hungary	Public, local authorities, government ministries and private insurers	Interviews, a public survey, a stakeholder workshop	No
[21]	2013	Ex-ante framework for flood damage assessment	Vietnam	Expert	Multi-criteria perspective to enables stakeholder participation	No
[22]	2012	Supporting tool for planning and developing an Integrated Flood Management for river management	Malaysia	Government department stakeholders	Stakeholder interests and expectations and how they could contribute toward an integrated approach	Yes
[23]	2012	Facilitate the decision making process in water related disasters	Thailand	Expert	Kendall's Coefficient of Concordance and the Kruskal Wallis H-test	No
[24]	2012	Impact assessment of management scenarios developed with active stakeholder's participation	Poland	Expert	Active stakeholder's participation	Yes
[25]	2003	The experiences of the Social and Institutional Responses to Climatic Change and Climatic Hazards (SIRCH) project	UK - The Netherlands - Spain	Seven cities in Europe and expert	Citizen Integrated Assessment (IA) Focus Groups. A stakeholder dialogue was conducted with sub-sets of experts	No
[26]	2000	Planning of regional water resources systems	USA	Expert and local stakeholder	Decision support system with which the decision maker can interact comfortably	No

Once framed the state of the art about the flood management and the stakeholder participation, a further classification has been developed in order to identify the level of participation and interaction. To define the level of participation it has been taken into account the categories of stakeholders engaged in the decision problem and in which phase of the process they participate, while for the level of interaction have been considered the methodology and techniques used for the interaction and if it has been conducted in group or individually. Figure 1 shows an outline of results obtained by classifying the set of papers resulted from the research.

Most of the papers are concentrated in the quarter with the lowest level of participation and interaction or in the quarter with the highest one. In fact, it is possible to underline from the review that: (i) stakeholders are involved in the whole decision process using interactive methodology or they are considered only for few phases, such as the criteria weight elicitation; (ii) they are interviewed individually and not in group; (iii) different methods of aggregation are used to sum up the multiple preferences.

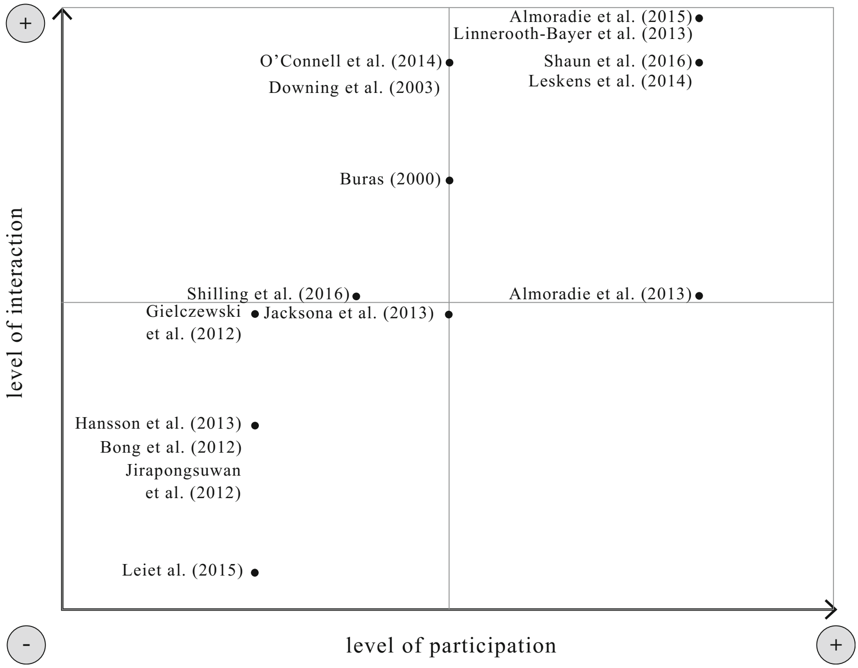


Fig. 1. Matrix of level of participation and interaction

3 The Role of Evaluation for Flood Risk Management

In order to define risk management plans and evaluating the consequent mitigation actions, a social integrated evaluation approach is proposed in order to integrate the different kind of stakeholders' involvement defined by the European Directive within a risk management planning process (Fig. 2).

In the proposed evaluation framework integrations are meant as a dynamic interaction between the risk management planning process and the evaluation process. Furthermore, in such an evaluation framework, the Stakeholder Analysis [27] can support the development of alternative scenarios for the mitigation of flood risk by structural and not structural actions with respect to local objectives, needs and values. More in deep, the Stakeholders Analysis plays a twofold role: on the one hand, the aim is to identify an "Actor Map" that performs relevant actions and has specific interests in the decision-making process in order to expose their goals and resources with respect to the project's realization. On the other hand, the objective is to integrate the "expert" knowledge with a "common knowledge" to support the construction of socio-economic strategic scenarios and the design choices.

According to Simon, [28] and Sharifi and Rodriguez, [29], the model is based on a flow of activities from analysis to implementation and monitoring, passing through choice, revision and decision phases. The framework represented in Fig. 2 highlights how each phase of the decision-making process involves the contribution of several evaluation methods and tools.

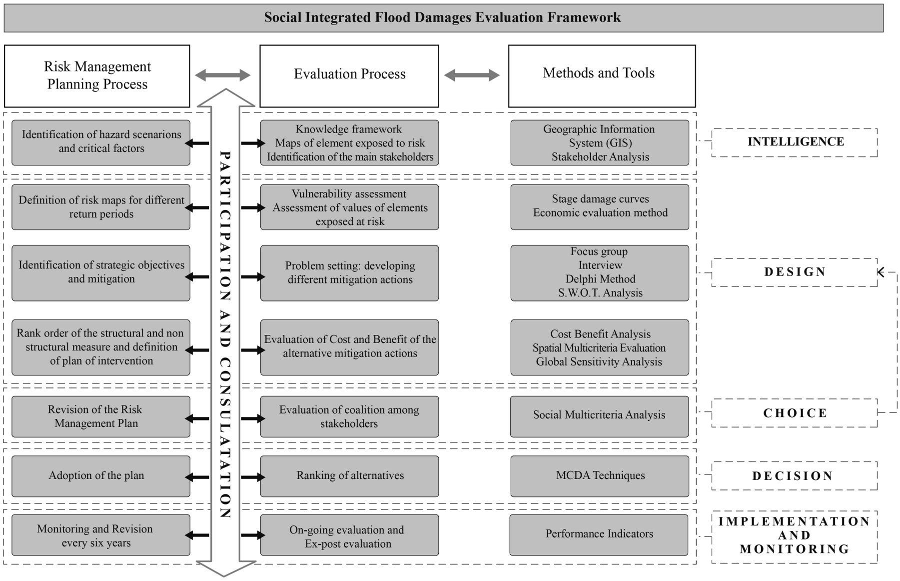


Fig. 2. The social integrated evaluation framework

4 Results and Conclusions

The analysis of the literature has allowed to obtain a general picture of the current level of stakeholders' participation and interaction, in addition to the operational tools and techniques generally used to support their engagement into flood management processes. Starting from the year of each contribution it is clear how the interest on this topic is quite recent. It is also important to underline how there are no Italian scholars active on this specific field of research, or more in general there are no case studies located in Italy.

Learning from literature review and trying to outline operational recommendations, a crucial issue to be faced is how to involve stakeholders in the different phases of the decision problem [30, 31], from the definition of their needs and expectations, with the aim of drawing possible strategies of action, until the selection of the most suitable alternative to solve it. Engaging them from the very first stage to the last one allows to frame a transparent and interactive evaluation process, flexible to changes according to potential suggestions and feedback received during the interaction. In order to achieve a general consensus, the analysis highlights also the importance to consider not only political and administrative actors, but experts with a specific knowledge in the field of urban water management, emergency management, urban planning or more in general environmental management [32]. Moreover, according to the participatory paradigm, another relevant category to involve is represented by people that are personally affected by the flood risk, such as people living nearby flood-prone areas, since they could give their own perspective and add a social dimension to the decision problem investigated by this research.

Due to the complexity of flood management, a social integrated multidimensional evaluation approach should be proposed in order to identify values, stimulate strategic thinking and enable the participation of all the stakeholders involved.

References

1. Brito, A.J., de Almeida, A.T.: Multi-attribute risk assessment for risk ranking of natural gas pipelines. Reliab. Eng. Syst. Saf. **94**(2), 187–198 (2009)
2. Torrieri, F., Batà, A.: Spatial multi-criteria decision support system and strategic environmental assessment: a case study. Buildings **7**(4), 96 (2017)
3. Freeman, R.E.: Strategic Management: A Stakeholder Approach. Pittman Books Limited, London (1984)
4. Freeman, R.E., McVea, J.: A stakeholder approach to strategic management (2001)
5. Harrison, J.S., St. John, C.H.: Strategic Management of Organizations and Stakeholders. West Publishing, London (1994)
6. Dente, B.: Understanding Policy Decisions. SAST. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-02520-9
7. Maskrey, S.A., Mount, N.J., Thorne, C.R., Dryden, I.: Participatory modelling for stakeholder involvement in the development of flood risk management intervention options. Environ. Model Softw. **82**, 275–294 (2016)
8. Brown, K., Adger, W.N., Tompkins, E., Bacon, P., Shim, D., Young, K.: Trade-off analysis for marine protected area management. Ecol. Econ. **37**(3), 417–434 (2001)
9. Rockloff, S.F., Lockie, S.: Democratization of coastal zone decision making for indigenous Australians: insights from stakeholder analysis. Coast. Manag. **34**(3), 251–266 (2006)
10. Nordström, E.M., Eriksson, L.O., Öhman, K.: Integrating multiple criteria decision analysis in participatory forest planning: experience from a case study in northern Sweden. For. Policy Econ. **12**(8), 562–574 (2010)
11. Lim, S.L., Bentley, P.J.: Evolving relationships between social networks and stakeholder involvement in software projects. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, pp. 1899–1906 (2011, July)
12. Aburto, J.A., Gaymer, C.F., Cundill, G.: Towards local governance of marine resources and ecosystems on Easter Island. Aquat. Conserv.: Mar. Freshw. Ecosyst. **27**(2), 353–371 (2017)
13. Shilling, F.M., Vandever, J., May, K., Gerhard, I., Bregoff, R.: Adaptive planning for transportation corridors threatened by sea level rise. Transp. Res. Rec.: J. Transp. Res. Board **2599**, 9–16 (2016)
14. Lei, Y., Liu, C., Zhang, L., Wan, J., Li, D., Yue, Q., Guo, Y.: Adaptive governance to typhoon disasters for coastal sustainability: a case study in Guangdong, China. Environ. Sci. Policy **54**, 281–286 (2015)
15. Almoradie, A., Cortes, V.J., Jonoski, A.: Web-based stakeholder collaboration in flood risk management. J. Flood Risk Manag. **8**(1), 19–38 (2015)
16. O'Connell, P.E., O'Donnell, G.: Towards modelling flood protection investment as a coupled human and natural system. Hydrol. Earth Syst. Sci. **18**(1), 155 (2014)
17. Leskens, J.G., Brugnach, M., Hoekstra, A.Y.: Application of an interactive water simulation model in urban water management: a case study in Amsterdam. Water Sci. Technol. **70**(11), 1729–1739 (2014)
18. Almoradie, A., Jonoski, A., Stoica, F., Solomatine, D., Popescu, I.: Web-based flood information system: case study of Somesul Mare, Romania. Environ. Eng. Manag. J. (EEMJ) **12**(5), 1065–1070 (2013)

19. Jackson, B., Pagella, T., Sinclair, F., Orellana, B., Henshaw, A., Reynolds, B., Mcintyre, N., Wheater, H., Eycott, A.: Polyscape: a GIS mapping framework providing efficient and spatially explicit landscape-scale valuation of multiple ecosystem services. Landscape Urban Plann. **112**, 74–88 (2013)
20. Linnerooth-Bayer, J., Vári, A., Brouwers, L.: Designing a flood management and insurance system in Hungary: a model-based stakeholder approach. In: Amendola, A., Ermolieva, T., Linnerooth-Bayer, J., Mechler, R. (eds.) Integrated Catastrophe Risk Modeling, pp. 199–216. Springer, Netherlands (2013). https://doi.org/10.1007/978-94-007-2226-2_12
21. Hansson, K., Danielson, M., Ekenberg, L., Buurman, J.: Multiple Criteria Decision Making for Flood Risk Management. In: Amendola, A., Ermolieva, T., Linnerooth-Bayer, J., Mechler, R. (eds.) Integrated Catastrophe Risk Modeling, pp. 53–72. Springer, Netherlands (2013). https://doi.org/10.1007/978-94-007-2226-2_4
22. Bong, C.H.J., Mah, D.Y.S., Putuhena, F.J., Bustami, R.A.: Tools for integrated river flood management (hydraulics modeling and logical framework analysis). ASM Sci. J. **6**(1), 47–60 (2012)
23. Jirapongsuwan, A., Englande, A.J., Fos, P.J.: Development of a water-related disaster decision model for nurses in Thailand. J. Med. Assoc. Thai. Chotmaihet Thangphaet. **95**, S120–S126 (2012)
24. Gielczewski, M., Piniewski, M., Kardel, I., Okruszko, T.: Application of SWAT for modeling management scenarios in a large lowland watershed in Poland. In: 21st Century Watershed Technology: Improving Water Quality and Environment Conference Proceedings, Bari, Italy, 27 May–1 June 2012, p. 1. American Society of Agricultural and Biological Engineers (2012)
25. Downing, T.E., Bakker, K., Lonsdale, K., Summerton, N., Syngedouw, E., Giansante, C.: Expert Stakeholder Participation in the Thames Region. Public Participation in Sustainability Science, pp. 187–200. Cambridge University Press, Cambridge (2003)
26. Buras, N.: Building new water resources projects or managing existing systems? Water Int. **25**(1), 110–114 (2000)
27. Dente, B.: Drawing Lessons for Comparative Policy Analysis from the Study of the Evolution of Privacy Policy (2011)
28. Simon, H.A.: Theories of bounded rationality. Decis. Organ. **1**(1), 161–176 (1972)
29. Sharifi, M.A., Rodriguez, E.: Design and development of a planning support system for policy formulation in water resource rehabilitation. J. Hydroinform. **4**(3), 157–175 (2002)
30. Torrieri, F., Oppio, A.: The sustainable management of flood risk areas: Criticisms and future research perspectives, Green Energy and Techonology, Springer, under press
31. Sophronides P.: Decision Support Systems for Participatory Flood Risk and Disaster Management, PhD Thesis, Athens (2016)
32. Dell'Ovo, M., Frej, E.A., Oppio, A., Capolongo, S., Morais, D.C., de Almeida, A.T.: Multicriteria decision making for healthcare facilities location with visualization based on FITradeoff method. In: Linden, I., Liu, S., Colot, C. (eds.) ICDSST 2017. LNBIP, vol. 282, pp. 32–44. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57487-5_3

Big Data Analytics to Improve the Decision-Making Process in Public Safety: A Case Study in Northeast Brazil

Jean Gomes Turet(✉) and Ana Paula Cabral Seixas Costa

Management Engineering Department, Universidade Federal de Pernambuco, Recife, Brazil
jeanturet@gmail.com, apcabral@hotmail.com

Abstract. The concern about national security has increased over the years. The large number of crimes has brought a variety of serious problems to Brazil and other countries around the world. Therefore, the major challenge, especially in Brazil, faced by public safety is how best to analyze large amounts of data so as to identify the factors that influence how crimes evolve. Thus, this paper analyzes public safety in the northeast of Brazil and proposes a decision-making model based on Big Data Analytics. This model is a part of a framework that will support decision processes by identifying the most dangerous places based on correlating data on location and the number of crimes from a large volume of crime data.

Keywords: Public safety · Big data · Hadoop framework

1 Introduction

Concern about national security has increased over the years. The huge number of crimes has caused Brazil and other countries around the world a variety of serious problems. In this context the major challenge faced by public safety is how best to tackle analyzing vast amounts of data in order to identify the factors that influence the evolution of crimes. This is especially needed in Brazil [1]. Thus, investing in control by policing and in the technology that supports this are seen to be important issues in crime prevention [1, 2].

To have efficient control over actual and potential criminal activity by means of policing, public safety needs to address how best to analyze the growing volume of crime data accurately and efficiently [3]. This is one of the greatest problems and challenges faced by public safety authorities because analysing large volumes of data is a difficult task for several reasons which include how to process such data, concerns about computational performance and the cultural factors that influence these analyses. It is in the nature of public security that several analyses of very large volumes of data must be made since decisions are made in real time. In these contexts, what public safety authorities require are Big Data solutions and therefore ways to achieve this so that the data can be processed more efficiently and with faster results and quality.

By investing in big data solutions (such as Hadoop and Data mining) supported by a model or a methodology from a decision support system, public safety will have the ability to accurately characterize, detect, anticipate and ultimately prevent crimes based

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 76–87, 2018.
https://doi.org/10.1007/978-3-319-90315-6_7

on a thorough analysis of some of the main factors, such as the past behavioral pattern of crimes in a region (pattern extraction) and the type of crimes in that region [3, 4].

Based on that, this paper analyzes public safety in the northeast of Brazil and proposes a decision-making model based on Big Data Analytics. This model is a part of a scheme that will support the decision processes that seek to identify the places of greatest danger by correlating place and the number of crimes committed from a huge volume of crime data.

The sequence of this work is as follows. Section 2 provides related studies for Big Data and some of its applications. Section 3 introduces the concept of Big Data and Multicriteria Decision. Section 4 presents a framework for analyses large volumes of data in public security. Section 5 contains a case study, discussion and results about the framework. Finally, the conclusion presents the limitations of this study and highlights the insights this work has gained.

2 Related Studies

The importance of public safety turns the focus of researchers' attention towards the amount of data. In this context, big data and its tools can offer a variety of advantages. In addition, a number of models are available in the academic literature, as well as discussions on controlling public security control based on Big Data and Data Mining. [5] develops a model that can predict criminal behaviors from a spatial choice with data mining techniques. [6] creates a Decision Support System (DSS) that deploys police efficiently by taking into account the territory in order to reduce criminal acts. [7] provides a framework that can identify crimes from different criminal activities. This is an investigative study on using spatiotemporally tagged tweets to predict crime. [8] extends this prediction by estimating Twitter and Kernel density and includes information coming from Twitter (the social media network) within an intelligent DSS.

Another line of research that has been widely applied has focused on discussing public security control based on Big Data Mining and its challenges. [3, 9] started a discussion on the possibility and development of a pattern for constructing a public security management system from big data solutions. The authors show the difficulties which are then encountered such as, for example, when the data pattern is not uniform and there is a huge variety of information (text, audio, image, video…). This leads to a huge difficulty, namely how to obtain useful information.

As to the literature on big data, data mining and predictive policing, most studies emphasize the technical aspects of tools and discussed how these tools have been used in this context. According to [6], predictive policing is defined as the use of a huge amount of data obtained from a variety of sources to produce results that can anticipate and prevent crimes. Mathematical and technological tools are used. This is why a common understanding has developed from the literature regarding the importance of using technological tools for mining big data to prevent crime. However, the literature does not provide an intelligent DSS model integrated with *Big Data Mining* in order to offer effective predictive policing. However, some studies do provide solutions for predictive policing by using isolated and non-integrated technological tools (*Big Data*

Mining and Decision Support Systems), which can lead to errors of interpretation and analysis with lower accuracy rates.

These studies and others in the literature show the importance of Big Data tools for analyzing large volumes of data on public security.

3 Big Data Analytics Tools and Multicriteria Decision Process

A brief overview of big data analytics and the multicriteria decision model will be given in detail in this section.

3.1 Big Data and Big Data Analytics

Over the last few years, organizations have had to work with different amounts of data from different sources. Analyzing these data has become increasingly complex due to the fact that robust tools and high processing power are required. In this context, this section discusses important concepts that are needed in order to reach a fuller understanding of big data and its tools.

Over the last decades, the amount of data has increased globally and to work with large-volume is complex. Based on this, the term "Big Data" and its tools sprung up around the world because it is applied to datasets that have vast amounts of data (millions). Big Data according to the authors [11, 12] is a problem due to the fact that organizations need to constantly analyze huge amounts of data and to store them.

When compared with the traditional databases, Big Data has five main V's: volume, variety, velocity, veracity and value [12–14].

On the other hand, the Big Data Analytics according to [10] is analytical techniques that are applied to big data. In this context, Big Data Analytics is the analytical tools and intelligent work undertaken on large, structured or unstructured data that is collected, stored and interpreted by very high-performance software. It involves cross-checking a near-boundless array of data from the internal and external environment [10].

The analyses of huge amount of data can be undertaken by many tools, such as HADOOP, MapReduce and OpenRefine. The Hadoop library (Fig. 1) is a framework that allows the distributed processing of large data sets through computer clusters using

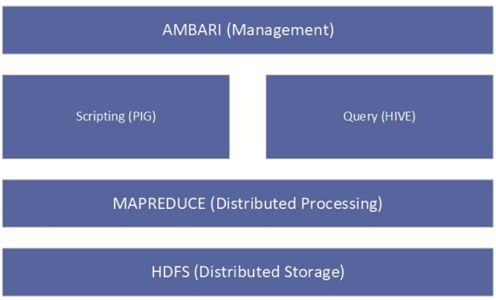


Fig. 1. Apache Hadoop ecosystem.

programming models [15]. It is designed to ensure broad scalability from a single server to a cluster with thousands of machines, each offering local computing and storage capacity. Instead of relying on hardware to provide greater availability, the library itself is designed to detect and address application layer failures so as to provide a high availability service based on a grid of computers.

MapReduce is a set of libraries that allow the processing of large amounts of data in parallel using all the hardware available in the Hadoop cluster. The MapReduce programming model consists of constructing a program formed by two basic operations: map and reduce. As for OpenRefine, it cleans data and transforms them into other forms [15, 16].

3.2 Multicriteria Decision Model

MCDA (multi-criteria decision analysis) methods are a collection of methods that consider criteria so as to support people or groups explore decisions [23]. These methods can be classified into three families: Single Synthesis Criterion, Outranking, and Interactive [17–20]. The single-criterion synthesis have a compensatory characteristic, use constant of scales. MAUT and AHP are examples of families of methods [21] On the other hand, in Outranking Methods there is no compensation between the criteria. They use weights (degrees of importance). The methods of the ELECTRE and PROMETHEE families are examples of Outranking Methods [17–19]. Finally, as for interactive methods, they are methods that alternate between stages of calculations, and produce a continuum of solutions, which provide immediate information about the preferences of the decision maker (DM) [19].

PROMETHEE II is an outranking method which aims to provide a complete preorder between alternatives. Initially, an outranking relationship is constructed, where a pairwise comparison between the alternatives for a given set of criteria is performed, in order to represent the DM's preference. Also in this step, the weights of each criterion (degrees of importance) are established. In the next step we need to explore the relation of outranking.

According to [22], the preference intensity of an alternative a with an alternative b is established according to how the DM's preference increases with the difference between the performance of the alternatives for each criterion $\left[g_i(\mathrm{a}) - g_i(\mathrm{b})\right]$ and assumes values between 0 and 1.

With the conclusion of the evaluations of the alternatives, two preorders are formed, the first being when the alternative "a" over-classifies all other alternatives and the second, when the alternative "a" is overclassified by the others (Eqs. 1 and 2).

$$\Phi^{+}(a) = \frac{1}{n-1} \sum\nolimits_{b \in \mathrm{A}} \pi(a, b) \tag{1}$$

$$\Phi^{-}(a) = \frac{1}{n-1} \sum\nolimits_{b \in \mathrm{A}} \pi(b, a) \tag{2}$$

In this context there are two flows: the first with the order of overcounting one alternative over the others, and the second, the first alternative being overclassified by the others. From this, we have the net flow for the ordering of the actions, as represented in Eq. 3.

$$\Phi(a) = \Phi^{+}(a) - \Phi^{-}(a) \tag{3}$$

4 Framework and Multicriteria Decision Model Proposed

The proposed framework comprises three main steps: the first is the acquisition of data, the second is the process of data management and the third is the decision-making process (Fig. 2). A brief overview of the framework and multicriteria decision model is given initially and then each component of the proposed framework and model will be explained in detail.

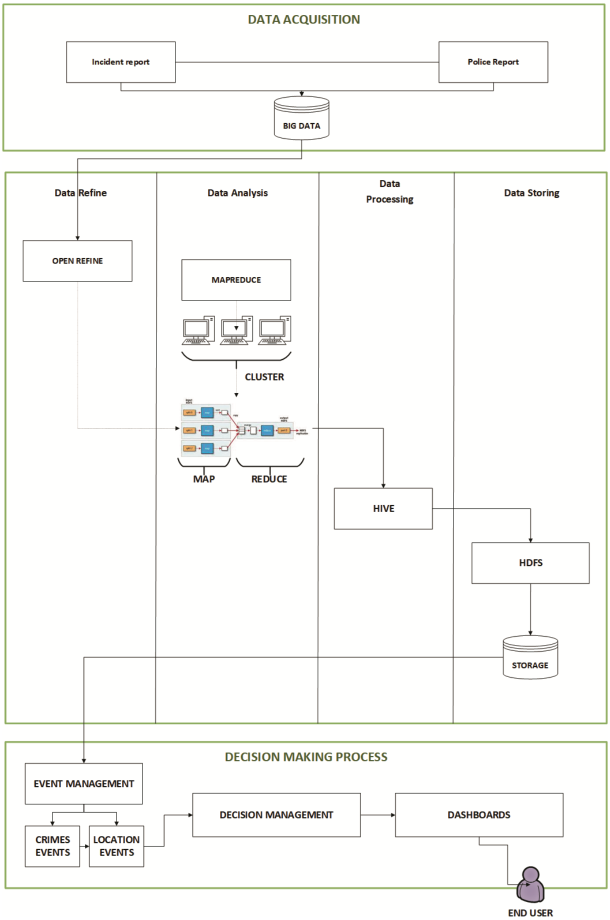


Fig. 2. Framework proposed.

4.1 Overview

Public security brings challenges for governments and society. Police departments must constantly deal with large amounts of data that happen in real time. In Brazil it is no different. The police department studied in this paper is located in the northeast of the Brazil where there is a constant need for actions that aim to reduce the number of crimes in the locality. In this context, this model is based on a real scenario, its purpose being to support the decision-making process by analysing large volumes of data. With this model, the police department and the social defence department will be able to identify areas with the highest numbers of crimes, and the relationship between crime and locality. It uses visual features (such as graphics) to make the best possible decisions regarding actions that must be carried out to minimize crimes in localities affected by violence.

4.2 Data Acquisition and Data Management

The acquisition of data basically takes place in real time. Police departments, the objects of study of this paper, use systems that aim to make the process of data capture easier. For example, the report of the occurrence of a crime. There is already an online system that allows people to access and register such reports. The data is stored in a database and it is available for the integrated access of the civil and military police, and the Social Defence Secretary and their support teams.

Another form of data capture takes place in each specialized military police battalion, where police officers input the notification of a crime reports and these data are then made available to civil and military police stations and barracks, and the Social Defence Secretary and their support teams.

The big challenge is to ensure the quality of the data, which does not happen consistently.

The management of data is the mediator between data acquisition and the application level. The components of this step are:

1. **Data Refining** has an important role in data analysis. The data quality aspect becomes a major challenge faced by public security. In this context, the second part of the model begins by applying a refinement algorithm, called OpenRefine (as shown in Sect. 3).
2. **Data Analysing** begin with 2 steps, a Map, which is the mapping and validation of the data and the another called Reduce, the input for which is the result of the previous Map phase, from which the final result is generated.
3. **Data Processing** uses HBASE and HIVE. HBASE is used to enhance the processing speed on Hadoop as it offers real-time lookups, in-memory caching, and server-side programming. HIVE uses a language called HiveQL (Hive Query Language), which transforms SQL statements into MapReduce Jobs running on the Hadoop cluster.
4. **Data Store** has the HDFS (Hadoop Distributed File System). This is a distributed file system, designed to store very large files, with standard access to data streaming, using server clusters easily found in the market, and is of low or medium cost. The HDFS has 2 types of Nodes: Master (or Namenode) and Worker (or Datanode). The Master stores

information about the distribution of files and metadata. Worker stores the data themselves. Finally, the derived intelligent decisions are transferred to the application level (decision-making process).

4.3 Multicriteria Decision Model

After analysing and identifying which neighbourhood has the highest crime rate and the most committed type of crime, it is important to measure a set of actions that aim to show the crime rates of the neighbourhood in question in a better way. In this context, in the last step of the framework, there is a multicriteria decision support model. In order to apply this model, an analyst is needed to support the DM. The model has six main steps as shown below:

Step 1: Identify neighbourhood with the highest number of crimes. In this step, the locality will be identified that has the highest number of crimes among those analysed. This part is checked by the first and second part of the proposed framework, where there is a large data analysis; **Step 2:** Identify Types of Crime. While identifying a locality with greatest number of crimes, this step also identifies the types of crime that have taken place and their frequency occurrence. This step is also identified by the framework; **Step 3:** Identify alternatives. Identifies alternatives that will lead to a greater reduction in crime in the neighborhood; **Step 4:** Set criteria. Selection of the criteria that will be used to analyze alternatives; **Step 5:** Apply the method chosen. Application of the most appropriate multicriteria method; **Step 6:** Results. Identify the actions needed and put them in order of priority. The results will be shown on Dashboards.

In the next section the applicability of the framework and the multicriteria model of decision support will be demonstrated by discussing a case study.

5 Case Study, Results and Discussions

As previously seen, one of the greatest challenges for public security in Brazil is how to tackle analyzing large volumes of data. During the day the public security systems receive several data that need to be dealt with in real time, since decision-making is constant. Several factors have a negative impact on the processing and analysis of such data, such as inadequate computational performance. In this context, a framework was proposed that supports the processing and analysis of data in a more efficient way and ensures that the process of decision-making is conducted with greater efficiency. A case study is used to demonstrate the efficacy of the method.

The case study was based on data made available in confidence, since these data directly affect society and safety issues. For this reason, all analysis will be based on coded names.

In order to analyze the neighborhoods with the highest rates of violence, there will be discretion regarding names, addresses, victims and precise locations of streets. The real inference is to show the applicability of the framework and the decision model when it handles large volumes of data. For this reason, the neighborhoods will be called

Neighborhood 1, Neighborhood 2, Neighborhood 3 and Neighborhood 4 so that the guarantee of confidentiality is respected.

The data were analyzed from a local database (MySql) on a computer with high processing power, content 2 TB of internal memory, 8 GB of RAM and a 4th generation CORE I5 processor.

In order to carry out this analysis, there was a connection with the Hadoop framework which established a cluster with three computers from the work environment of the Cloudera QuickStart version. This environment enables the integrated analysis of data from clustering (from Hadoop) to be sent to dashboards for visual analysis.

The first part of the framework proposed in this paper is concluded, after data has been collected from two main points: the report of the occurrence and the police source.

The second part of the framework consists of several main phases, the first being data refinement. Especially with regard to the data coming from the police being complete, there are, in practice, various inconsistencies, spaces not completed and gaps incorrectly filled in. For this reason, refining the data is important since this will improve the quality of the data without which there would be a negative impact on the final result. In this context, this refinement was made on the available data, thereby minimizing errors and improving the corresponding analyses (Fig. 3).

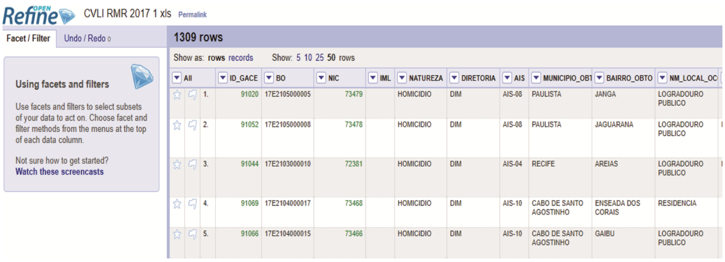
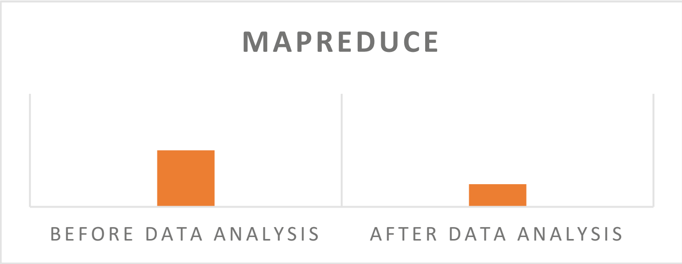


Fig. 3. Data refining.

After data refinement, the data are analyzed. In this step, Hadoop conducts the mapping and reduction process. The map operation receives a key/value pair and generates an intermediate set of data, also in key/value format. The Reduce operation is performed for each intermediate key, with all sets of intermediate values associated with that key combined.

After this step there was compression after approximately 30% (as shown in Graphic 1) in the number of lines associated with the database. The quality of the data is maintained, since in this step, the only data that are grouped are those which have the same previously defined characteristics, thus reducing the number of available lines in the database.



Graphic 1. MapReduce.

After the data have been mapped and reduced, they need to be processed. In this step, HIVE is used (as explained in Sect. 3). In this moment, a query (a standard declarative search language for relational database) can be used to identify which neighborhood has the highest number of occurrences of crime. By doing this, neighborhood number 2 was the one with the largest number of reports on the occurrence of crimes. Use of relationship analysis showed that in this neighborhood there is a large number of homicides, for which there are several reasons, including the crime of armed robbery resulting in a death. One of the queries used to identify neighborhoods with the highest occurrence number and, consequently, the type of crime, is identified in Fig. 4:

```
SELECT TOP(5) MUNICIPIO_OBTO, COUNT(BO) FROM CVLI
GROUP BY MUNICIPIO_OBTO
ORDER BY n DESC
```

Fig. 4. Query code.

Finally, this data is stored in a HADOOP HDFS mechanism, which is a distributed file system, designed to store very large files with standard access to streaming data, using server clusters easily found in the market at low or medium cost. This stored data is analyzed later.

To conclude this section, it is verified that neighborhood 2 has the highest number of crimes, largely consisting of homicides.

This result shows the need to determine a set of actions that will improve security in this locality.

5.1 Multicriteria Decision Model

The third part of the proposed framework is a multicriteria decision model. This model is an integral part of the framework and it sets out to identify a set of actions that will bring improvements in security to the neighbourhood identified in the previous analysis. This neighbourhood (neighbourhood 4) requires actions to reduce the number of crimes. However, an appropriately qualified and experienced analyst is needed to confirm what actions should be carried out and what the order of priority should be in order that the planning for public security is conducted correctly. The scenario for this is based on realistic provincial data and the case study.

The PROMETHEE II method was used in this paper mainly for two reasons:

- The method has a non-compensatory characteristic, where a weight (a degree of importance) is used for each criterion. For this study, the notion of weight is the most adequate, since there is no compensatory characteristic between the criteria (constant of scales);
- The method is used for the ordering problem and was adapted for the purposes of this paper. The purpose of the model to be presented is to order a set of actions, and to prioritize them so that this may meet the challenge the police and authorities face by resulting in improvements in public safety.

Table 1 shows the alternatives that could be adopted to minimize crime rates in the locality in question (neighbourhood 4). Having identified the needs of the region and the DM's preference (the Secretary for Public Security of the region studied), the main alternatives were identified as were the criteria for the analysis of these. The Secretary for Public Security has appointed a liaison officer who will also participate in the decision-making process.

Table 1. Alternatives.

	Alternatives	Summary description
A1	Expansion of the number of police vehicles in the locality	Increase the number of police vehicles in the locality to minimize robberies
A2	Integrated routing system	Establish an information system that automatically routes police vehicles by using real-time data
A3	Increase in police force in the locality	Increase police force in the locality
A4	Set up a local police station	Set up a local police station to increase security

Table 2 shows the criteria that will be used to evaluate alternatives and gives a brief description of these.

Table 2. Criteria.

	Criteria	Criteria scales
C1	Investment cost	Monetary (Brazilian Reais)
C2	Maintenance cost	Monetary (Brazilian Reais)
C3	Time taken to implement action	Days
C4	Impact on public safety in the locality	3-point qualitative scale

In order to identify the weights, the DM/client was able to determine the most appropriate values for the representativeness of the importance of the criteria, since the high degree of alternatives and criteria do not cause the model problems (Table 3). The preference function for the criteria was the usual (Table 4).

Table 3. Evaluation matrix.

	C1	C2	C3	C4
Min/max	Min	Min	Min	Max
Weight	0.15	0.10	0.18	0.30
A1	500,000	35,000	60	Good
A2	250,000	10,000	200	Average
A3	100,000	45,000	40	Good
A4	350,000	40,000	150	Good

Table 4. Final ranking.

Alternatives	**Φ**
A3	0.4521
A1	0.0594
A4	−0.0594
A2	−0.4521

The multicriteria model proposed identified that alternative A3 (Increase of the police force) was in the first position (Table 3). In fact, increasing the number of police officers may well reduce the criminality rate. Police officers present on streets lead to security being improved and the population to sense an increase in public safety.

A sensitivity analysis with variation of criteria weights of ±10% was performed and there was no change in the order between the alternatives.

6 Conclusions

One of the greatest problems and challenges faced by public safety is how best to analyze large volumes of data, since regions with a higher level of crime must be constantly identified in order to make better decisions. In the northeast of Brazil, it is no different. The police need to make constant analyses in order to take the best action to address a particular problem.

Thus, this paper put forward a framework, the purpose of which is to analyze large volumes of data. For this, the tools and techniques of big data analytics were used with a view to aiding the process of analysis and improving performance. In this context, it was verified that the framework is efficient since it succeeded in identifying, from a high volume of data, the neighborhood that has the highest number of monthly crimes. Secondly, a multicriteria model of decision support was successfully used to identify a set of actions, and to order these actions so as to bring about improvements in public safety in this neighborhood.

Regarding the limitations of the work, this study is more restricted to the aspect of the neighborhood with the highest level of violence and was not extended to include studies of behavior and of the tendencies for certain types of crime to be committed. Other studies could usefully undertake analysis of these matters.

References

1. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. IEEE Comput. Soc. **37**, 50–56 (2004)
2. Seele, P.: Predictive sustainability control: a review assessing the potential to transfer big data driven ‘predictive policing’ to corporate sustainability management. J. Clean. Prod. **153**, 673–686 (2017)
3. Williams, M.L., Burnap, P., Sloan, L.: Crime sensing with big data: the affordances and limitations of using open-source communications to estimate crime patterns. Br. J. Criminol. **57**, 320–340 (2017)
4. Amoore, L., Raley, R.: Securing with algorithms: knowledge, decision, sovereignty. Secur. Dialogue **48**, 3–10 (2017)
5. Xue, Y., Brown, D.E.: Spatial analysis with preference specification of latent decision makers for criminal event prediction. Decis. Support Syst. **41**, 560–573 (2006)
6. Camacho-Collados, M., Liberatore, F.: A decision support system for predictive police patrolling. Decis. Support Syst. **75**, 27–37 (2015)
7. Li, S.-T., Kuo, S.-C., Tsai, F.-C.: An intelligent decision-support model using FSOM and rule extraction for crime prevention. Expert Syst. Appl. **37**, 7108–7119 (2010)
8. Gerber, M.S.: Predicting crime using Twitter and kernel density estimation. Decis. Support Syst. **61**(1), 115–125 (2014)
9. Gong, X., Sun, J., Zhang, Y.: The discussion of the possibility and development pattern for constructing the public security management system based on big data. In: International Conference on Materials Engineering and Information Technology Applications, MEITA (2015)
10. Elgendy, N., Elragal, A.: Big data analytics in support of the decision-making process. Proc. Comput. Sci. **100**, 1071–1084 (2016)
11. Renu, R.S., Mocko, G., Koneru, A.: Use of big data and knowledge discovery to create data backbones for decision support systems. Proc. Comput. Sci. **20**, 446–453 (2013)
12. Grillenberger, A., Fau, F.E.: Big data and data management: a topic for secondary computing education. In: Proceedings of the 10th Annual International Conference on International Computing Education Research, ICER 2014, pp. 147–148 (2014)
13. Chang, R.M., Kauffman, R.J., Kwon, Y.: Understanding the paradigm shift to computational social science in the presence of big data. Decis. Support Syst. **63**, 67–80 (2014)
14. Dobre, C., Xhafa, F.: Intelligent services for big data science. Future Gener. Comput. Syst. **37**, 267–281 (2014)
15. White, T.: Hadoop: The Definitive Guide, 4th edn. O’Reilly Media, Newton (2015)
16. Verborgh, R.: Using OpenRefine. Packt Publishing, Birmingham (2016)
17. Roy, B.: Multicriteria Methodology Goes Decision Aiding. Kluwer Academic Publishers, Dordrecht (1996)
18. Sadovykh, V., Sundaram, D., Piramuthu, S.: Do online social networks support decision-making? Decis. Support Syst. **70**, 15–30 (2015)
19. Vincke, P.: Multicriteria Decision-Aid. Wiley, Hoboken (1992)
20. Pardalos, P.M., Siskos, Y., Zopounidis, C.: Advances in Multicriteria Analysis. Kluwer Academic Publishers, Dordrecht (1995)
21. Kenney, R.L., Raiffa, H.: Decision with Multiple Objectives: Preferences and Value Trade-Offs. Wiley, New York (1976)
22. De Almeida, A.T.: Processo de decisão nas organizações: construindo modelos de decisão multicritério. Atlas, São Paulo (2013)
23. Belton, V., Stewart, T.: Multiple Criteria Decision Analysis: An Integrated Approach, 1st edn. Springer, Heidelberg (2002). https://doi.org/10.1007/978-1-4615-1495-4

An Interactive Learning Environment Based on System Dynamics Methodology for Sustainable Mobility Challenges Communication & Citizens' Engagement

Glykeria Myrovali[1(✉)], Georgios Tsaples[2], Maria Morfoulaki[1], Georgia Aifadopoulou[1], and Jason Papathanasiou[2]

[1] Hellenic Institute of Transport/Centre for Research and Technology Hellas (HIT/CERTH), 6th km Charilaou-Thermi Road, 57001 Thessaloniki, Greece
{myrovali,marmor,gea}@certh.gr

[2] Department of Business Administration, University of Macedonia, Thessaloniki, Greece
gtsaples@gmail.com, jasonp@uom.gr

Abstract. Serving the goal of enhancing the participatory approach of sustainable urban mobility planning for delivering acceptable and viable mobility plans, the current paper presents the MOTIVATE Interactive Learning Environment (ILE)/Game. Based on the System Dynamics (SD) methodology and answering to the need for catching up to the interactivity trend, the MOTIVATE ILE offers the user with a simplified experiential procedure for understanding the consequences of mode choice and sustainable decision making. Moreover, the rewarding system proposed for allowing the performance of actions while using the ILE transforms the user into an active agent of mobility planning by asking him/her to provide travel data and opinions for the improvement of city's daily transportation performance.

Keywords: Sustainable mobility · Interactive learning environments
Serious gaming · Awareness raising · Behavior change

1 Introduction

Given that about 52% of the world population and over 70% of EU citizens live and perform their daily trips in urban areas, the identification of cost-effective ways in managing urban mobility naturally becomes one of the first challenges and, in parallel, flagships at cities agendas. Sustainable mobility measures are selected and tailored by city planners so as to meet specific needs. Broad dimensions of sustainable mobility planning include promotion of Public Transport (PuT), alternative modes (walking, cycling) and alternative fuel driven vehicles (e.g. electric), while in order to achieve the desired modal shift, strong emphasis is given on multimodality enhancement (e.g. development of fully interconnected and accessible terminals), on Intelligent Transport Systems (ITS) and Information & Communications Technology (ICT) development and on services integration (integrated information, integrated ticketing and harmonized

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 88–99, 2018.
https://doi.org/10.1007/978-3-319-90315-6_8

timetables among PuT modes) [1–8]. Mobility measures strategic assessment is widely undertaken in order to conclude in the formulation of effective mobility action plans and in the best possible funding sources allocation [3, 9] while transferability analysis identifies transferability principles [10–13].

Although the effectiveness of mobility management is undoubtable, what is still in its infancy is the participatory approach on the mobility planning [14, 15]. Transforming todays' cities into sustainable urban environments meeting global sustainability goals highly depends on citizens' collective endeavor and on a deep and radical change in (travel) behavior [16–20].

Based on the above and taking into account that real transition can be made only from within, the current paper is built on the need to invest in practical and effective awareness raising methods for transforming citizens into active agents of change. The channel to persuade and engage citizens' into mobility decision making described in the paper is an Interactive Learning Environment [20] in the form of an abstract serious game [21–23]. The paper reports on the main output of MOTIVATE (Interreg-MED) project, the MOTIVATE e-platform [24].

The remainder of the text is organized as follows: Sect. 2 reviews ICT based educational approaches in citizens' engagement in (mobility) planning and in awareness raising; Sect. 3 provides a summary of the MOTIVATE e-platform, part of which is the presented ILE; Sect. 4 presents the methodological approach on the simulation model and the interactive learning environment and further describes SD theory & system thinking; Sect. 5 concludes on key messages and considerations for improvements and content enlargement of the presented interactive learning environment and its applications.

2 Exploitation of ICT-Based Enabling Tools for Citizens' Engagement in Mobility Planning

Sustainable urban mobility era seems that can only be reached if innovative approaches that strongly engage citizens in the decision making and in the plans implementation phases apply. The term "innovative" here refers to the adoption of ICT based channels able to motivate citizens in being part of the transition era. ICT based tools for encouraging citizens' involvement in the sustainable planning procedure refer to, free of time and location constraints, technology-mediated forms of citizens' participation; no need for physical attendance at conventional stakeholders events where citizens are usually represented by just few members taking the role of simple observers [14, 15, 25].

2.1 Data Collection Enabling ICT Based Tools

Today, cell phones and GPS are useful enabling tools for traffic crowdsourced data collection, being simultaneously a cost-effective alternative to the traditional household or phone travel surveys. Data collection is based on a combination of proprietary probe-vehicle data, proprietary sensor data and public sensor data [26, 27]. Aggregated historical and real time traffic data (e.g. speeds, travel times) from contracted commercial providers (e.g. TomTom, Inrix, HERE, Cellint) collected with travelers' consent is an important asset for transport planners; the big traffic data era is a reality since years however what is still missing is the optimum translation of this enormous database to valuable and manageable knowledge.

Accident, event and general traffic related experience reporting apps are another way to collect traffic and city's performance information (e.g. Waze platform, Tiramisu Transit-vehicle Tracking App, Street Bump Smartphone App, SeeClickFix, Swiftly). Travel data can also be collected via social media which use increases significantly at daily base. Through social media, the users can "log", "post", "(re)tweet", "check in" and perform a variety of other "actions" which allow them to share experiences, thoughts, needs while at the same time provide input to a wide traffic database in an unstructured or semi-structured way [28–31]. Smartphone apps and websites are also suggested for traffic data collection projects [14, 31–33].

Another way to collect travel data that is proposed by the current paper is through the use of serious games and interactive learning environments. Travel data provision can be part of a serious game (embedded procedure) or linked to the rewarding system of a game (points earned for each trip diary completed in the game).

2.2 Travel Related Opinions and Suggestions Collection

Mobility goal setting, prioritization of interventions and optimal distribution of the limited public budget for achieving sustainability objectives, ask for citizens' participation in the strategic planning since only in this way acceptability and viability of measures are guaranteed. Serving this goal, social media and crowdsourcing apps are mobilized supporting traditional opinion capturing methods; rating polls, debate platforms, citizen forums & accounts on social media are only some forms of innovative ways to open a constant dialogue and communication with citizens. Building relations though online social networking and apps, although in early stages, has been proven effective in terms of results quality and participation quantity [26, 34, 35].

2.3 Sustainable Mobility Awareness Raising

Serving the need to gain public confidence and acceptability for the proposed mobility measures [36], social media have started being widely used (e.g. top hashtags #biketoworkday, @GamesTravel2014 Twitter account). Persuasive sustainable mobility, as also in its generic form, refers to the achievement of a profound change of travel behavior respecting sustainability principles typically by raising individuals' awareness of their choices, behavior patterns and the consequences of their activities [37]. Smartphone and web apps aiming to influence travelers' choices in favor of alternative to private car options have arisen [38–40]. Serious games and Interactive Learning Environments have started also finding their position inside the awareness raising complex procedure [23].

3 The MOTIVATE E-Platform

Following the above mentioned ICT principles, the MOTIVATE (MED 2014-2020) project goes a step further and make practical proposals for the use of ICT enabling technologies in the development of SUMPs (sustainable urban mobility plans) and generally in the transportation planning. Unlike the traditional data collection methods, where residents or visitors are “passive” data sources, the innovative approach of MOTIVATE project, currently under development, lies on their active involvement in transport data collection/management, problems’ identification and mobility related measures’ evaluation.

The core result of MOTIVATE project is the development of a technological platform which will be used as a communication channel between authorities and residents/ visitors. The 4 services of the MOTIVATE platform are:

- Service No 1: “Trip Diaries”

Where user is asked to enter its daily trips at real time (GPS enabled – “start tracking”/ “stop tracking”) or after trip completion. Additional data for each trip are optional (e.g. mode of transport, trip purpose etc.).

- Service No 2: “Measures Evaluation”

The “Measures Evaluation” service aims to collect evaluations (in the form of rating) regarding the performance of existing mobility measures and of the existing transport network. User is asked to select city, measure to be evaluated and express for each criterion its overall star satisfaction.

- Service No 3: “Future interventions assessment”

In a similar approach to Service No2, the third service, the “Future interventions assessment” service aims to collect perceptions regarding the usefulness of future mobility interventions (in the form of rating).

- Service No 4: “The MOTIVATE GAME”

The points earned in the first 3 services can be used for performing actions in the MOTIVATE game/ILE. The game supports in this way the involvement of citizens in data and opinion collection services while simultaneously adds on awareness raising and deep behavior change.

4 Methodology

The aim of MOTIVATE game is to inform and teach users for sustainable mobility benefits. The main challenge of achieving sustainability goals is to persuade citizens for the long-term effects of their current choices. Furthermore, human behavior is always a complex issue to be approached and modelled; mobility choices include aspects not easily quantified i.e. safety, comfort and value of time are involved [41]. Moreover, urban mobility depends on the combination of many parts-individuals’ choices that which often interact and influence each other and on current political decisions. In this context, decision-making becomes dynamic since it requires several interdependent decisions [42, 43]. Finally, it is hard for people in general (and decision-makers in particular) to understand non-linearities and feedback loops and how they can be manifested in real-world systems [44]. As a result, for the model to be understandable and usable by non-technical users, it must offer a limited amount of interactions without losing the main information and aspects of transportation complexity, while at the same time it must offer to users the possibility to arrive at their own conclusions regarding the issues of sustainability and complexity.

For these reasons, simulation, a computer-based representation of real-world system offering great insights regarding system's behavior, comes as a natural candidate for the development of such a model. SD [45, 46] is a computer-based methodology that makes use of integral equations with the purpose of studying the behavior of systems over time. Hence, it provides insights to decision-makers on how a policy might change the state of the system under study. The modeling tools of SD are stocks, flows, feedback loops and time delays that give rise to dynamic complexity, inherent in transportation systems, through quantitative simulations [46]. Furthermore, it allows the incorporation of "soft" variables, typical of transportation systems, which generally are not taken into consideration in the traditional econometric models. Behind the use of SD lies the idea that a system's structure defines its behavior [46] thus, by analyzing and representing adequately the relations between its various elements, it could be possible to define accurately the structure of the problem under study, which would ultimately lead to an increased understanding of its behavior over time.

Integrating the various aspects retrieved through the theoretical framework, SD simulations have the power to represent the main mutual influences of the various parameters, defining for each influence if it is positive or negative, thus defining also the value and timing of these influences. This holistic approach requires the consideration of the entire context and on the other hand, parameters/system elements that are considered weak or not related to the processes of interest to be omitted, with the purpose of avoiding to end up with a model that is difficult to manage and/or interpret. Finally, even an elaborate SD model can be embedded in an Interactive Learning Environment, which could ease the communication of results and system's behavior to end-users.

4.1 Model Structure

The game illustrates the complexity of urban transportation systems and at the same time it is simple enough to communicate it to the users of the application. Under this spectrum, the game offers a limited number of actions able to be performed by the users, so that it will not overload him/her with information and become disproportionally difficult and thus unattractive.

To achieve the objectives of the game, restricted by the aforementioned characteristics, a simulation model is developed with the methodology of SD. The context is that of an urban environment with various modes of transport and the simulated population performing trips by choosing mode based on several criteria over a period of simulated time. The purpose for the user is to achieve the highest score on main KPIs (Key Performance Indicators).

The following diagram represents the main causal relations among the variables that determine the mechanism of the game and generate its behavior. The variables in black are the main model variables that are hidden from the user; variables in blue represent the potential actions that the user can perform (assuming that he/she has the required points) and the variables in red represent the KPIs that will be the results of the actions performed. The arrows demonstrate a causal relation between two variable, while the signs + and − show the direction of that relationship. The + sign illustrates that the direction of the change is similar for both the variables and the − sign that the direction of the change is opposite between the two variables. Thus, the variable traffic KPI is affected by two variables: Private transport users and maximum car capacity. The direction of change is opposite for the two cause-variables, thus they have a counteracting effect on the traffic KPI. The final behavior of the traffic KPI will depend on the values of the two variables that in turn depend on the whole state of the model. These causal relations in the model form several feedback loops. A feedback loop means that a change in one variable will generate changes in several other variables which in turn will result in a new change in the variable that initiated the loop (Fig. 1).

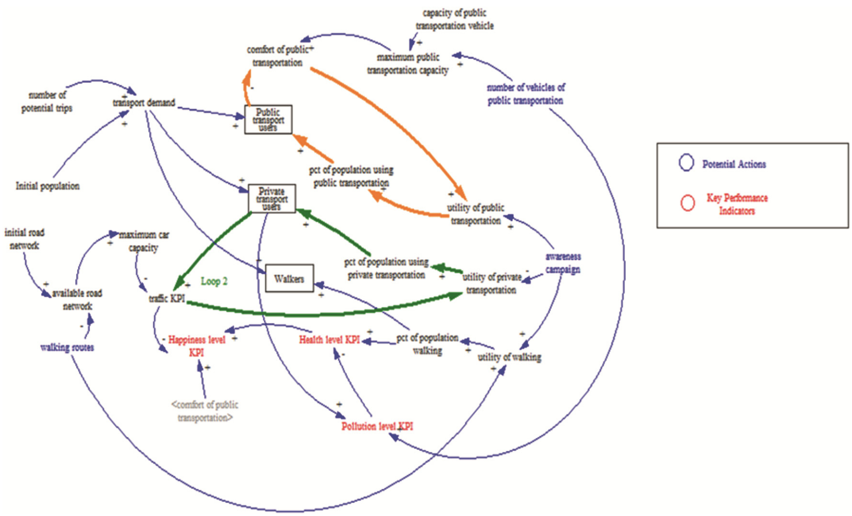


Fig. 1. The causal loop diagram (with the main loops in orange and green) (Color figure online)

For example, a loop is formed among the variables: utility of PuT →+ pct of population using PuT →+ PuT users →− comfort of PuT →+ utility of PuT. The explanation of the loop is the following: Let's assume an increase in the utility of PuT. Then this increase will result in an increase in the percentage of people who use PuT. The particular increase will result in a decrease in the comfort of PuT (− sign) which will ultimately decrease the utility of PuT. Depending on the values of the variables, the last decrease may cancel the initial increase in the utility of PuT, thus generating different behavior on the whole system from the initially assumed.

Another loop is among the variables: utility of private car →+ pct of population using private transportation →+ Private transport users →+ traffic KPI →− utility of private transportation.

An increase in the utility of private car will result in an increase in the percentage of the population that use cars which in turn will result in an increase in the Private transport users. This will increase the traffic KPI which will ultimately decrease the utility of private transportation. Depending on the values of the variables, the last decrease may cancel the initial increase of the utility of private transportation thus resulting in a different system behavior than was originally assumed.

Apart from the loops among variables, the loops themselves interact with each other. Following in the previous example, there is an interaction between Loop 1 and 2. A final decrease in the utility of PuT (Loop 1) will result in an increase of the other utilities (including that of private transportation), thus activating the changes in Loop 2. The interaction between loops is not illustrated in the diagram for economy of space and clarity.

The user is able to act on three variables of the model (indicated with the blue color in the diagram) and can see the results of the variables that are indicated in red in the diagrams: Pollution level KPI, Health level KPI and Happiness level KPI.

As for the potential actions, the user can manipulate them and generate the new system state (KPIs). The values of the KPIs in turn will force the user to decide on another action in the game.

The conceptual definition of the three potential actions is the following:

- Awareness campaigns: It is a theoretical variable that cannot be easily quantified. In the model it takes values between 0 and 1, with 0 indicating that there is no campaign in action and 1 that the campaign is activated and in full intensity.
- Coverage of PuT (number of PuT vehicles): An increase in the number of available PuT vehicles/coverage that - in theory means - more people will be attracted to PuT, since it will be more comfortable to do so
- Number of pedestrian routes: It takes values between 0 and 50 km. The variable represents how much of the available road network will be transformed to pedestrian routes.

To perform one of the potential actions the user must have the appropriate number of points that can be gained by using the other 3 services of the MOTIVATE platform. The model of the game simulates for 48 months and the purpose of the user is at the end of the simulation time to achieve a high score on all three KPIs (maximize Happiness and Health Level and minimize Pollution level). At the end of the simulation time, a turn of the game is completed.

4.2 Indicative Results

There is a large combination of scenarios that involve the playing of the game. As an example, several scenarios will be displayed to provide more insights into how the users' actions affect the behavior of the system (Table 1).

Table 1. Scenarios that are simulated in the model of the game

A/A	Action in the game	Required points
1	None (initial conditions)	–
2	Increase vehicles of PuT by 1	2
3	Increase walking routes by 10 km	8
4	Activate awareness campaign and increase intensity by 0.1	9
5	All of the above	19

The results of the actions described in the table above are illustrated in the figure below (Fig. 2).

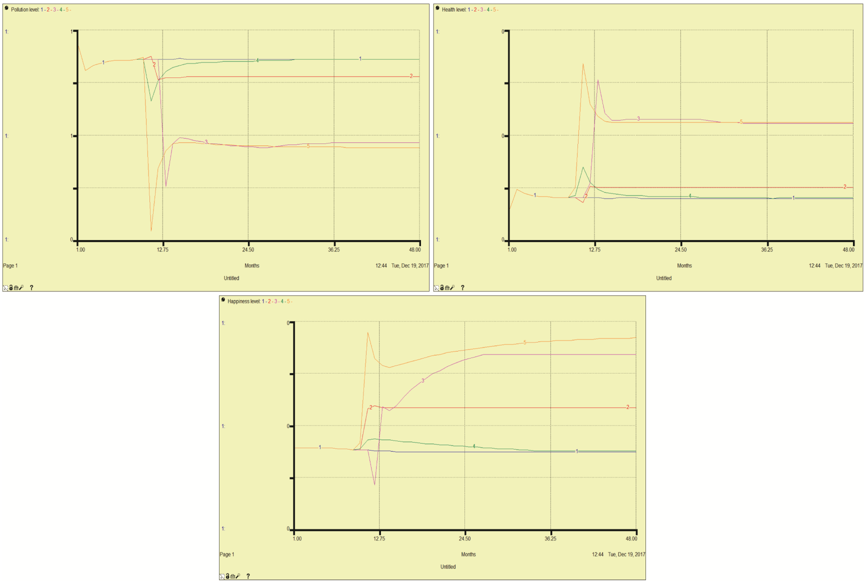


Fig. 2. Results of the three KPIs (health level on the left, happiness level on the right, pollution level below) based on the above scenarios

The array of scenarios demonstrates the effectiveness of the different sustainability measures. First, when no policy is simulated, the system is in equilibrium and since the private vehicle is the most preferred mode of transport (based on personal criteria and mainly on comfort), the city suffers from high pollution and low health and happiness. By increasing PuT services, more people leave their cars, thus all KPIs behave in the desired direction. However, there is a small period of time when the Health and Pollution KPIs have the opposite effects; the introduction of new vehicles in the system means that there will be larger emissions until the population gets accustomed to the new conditions and the system reaches equilibrium. Thus more emissions means that there will be a small decline in the health level (small decline in line 2 of the figures above) for a limited period of time. As a consequence, a sustainability intervention might produce undesirable effects in the short-term until the system reaches the new state of equilibrium and the favorable effects of the policy manifest.

The increase of walking paths also has positive effects to the KPIs. Similarly, there is a small decline in the Happiness KPI for a short period until the system reaches the new equilibrium and the travelers change behavior in favor of walking and walking-cycling alternatives. This is expected, since the construction of new walking paths means that the city network changes, the previous car users still (for a little) use their cars making more maneuvers for reaching their final destinations, which increases the traffic of the city thus, temporarily, decrease the happiness level of the population. However, as the system reaches the new equilibrium the full positive effects of the measure act on the KPIs on the desired direction.

The increase of the intensity of the awareness campaign has positive effects on the city, but only for a short period of time; awareness and sustainability training is a constant procedure in order to keep citizens' interest and willingness alive.

Finally, by combining all three policies the effect is the optimum one. As a result, the game illustrates that individual and fragmented actions can only have limited results, however, integrated planning and measures' combination can bring long lasting effects. The model (and the game) demonstrates emphatically that a more holistic approach to decision-making with regards to mobility is required. Decision-makers and users need to view the transportation system as a whole and design and act based on the effects and interactions of its various parts. Finally, the game offers insights to the users by placing them in the position of a decision-maker and demonstrating how difficult it is to achieve sustainability and how individual choices can affect the entire system.

5 Conclusions

The MOTIVATE GAME was designed in a rationale of providing an attractive training experience to users-travelers while engaging them indirectly in the mobility planning procedure; it illustrates in the most intuitive way the complexity of transportation system, it engages the user in a dummy decision-making process and provides insights into the effects that any decision might have to the behavior of the transportation system. From the authorities' side, it is estimated to add in their understanding of travelers' needs, therefore to formulate a useful arsenal for effective mobility plans design and implementation.

The model at the core of the game is a simulation model developed with the methodology of SD. Despite the fact that the model is not an accurate depiction of an urban transportation system, it conveys the main messages regarding its performance. The user can perform three actions to the game (pedestrian routes, PuT vehicles, awareness campaigns levels) only if he/she has the required points gained through the other three crowdsourcing based services of the platform (where data essential for the planning procedure are elicited).

The game can illustrate to the user that individual and fragmented actions do not have a significant result towards a more sustainable transportation system, because of its inherent non-linearities and complexity, however what is needed is a deep and long-lasting behavior change from the largest part of the citizens. Users are experiencing a web-based attractive training activity to adopt a top-down approach and view the transportation system as a sum of interrelated parts that affect not only themselves and each other but also the environment, the economy and the viability of the future. Finally, the game offers insights to its users, as it puts them in the position of a decision-maker, and demonstrates how difficult it is to achieve sustainability and how individual choices and trade-offs can affect balances of the entire system.

The model described has limitations; a balance between simplicity and correct information communication is tried to be achieved and future improvements are already being scheduled. The context of the MOTIVATE Game is already planned to be enlarged while the services of the MOTIVATE platform will be also enhanced and tailored to specific mobility needs so as to offer more opportunities to the city planners and the decision makers. Personalized differentiations along with personalized notifications supporting sustainability are planned as well. Finally, the game and the services will be presented in the future in an embedded version where a more advanced learning experience will be offered.

References

1. Litman, T.: Module 2b: mobility management. Victoria Transport Policy Institute, GTZ Transport and Mobility Group (2003)
2. Litman, T.: The online TDM encyclopedia: mobility management information gateway. Transp. Policy **10**(3), 245–249 (2003)
3. Meyer, M.D.: Demand management as an element of transportation policy: using carrots and sticks to influence travel behavior. Transp. Res. Part A: Policy Pract. **33**(7–8), 575–599 (1999)
4. Meyer, M.: A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility. Institute of Transportation Engineers, Washington (1997)
5. Corazza, M.V., Guida, U., Musso, A., Tozzi, M.: A European vision for more environmentally friendly buses. Transp. Res. Part D: Transp. Environ. **45**, 48–63 (2016)
6. Morfoulaki, M., Myrovali, G., Kotoula, N.: Increasing the attractiveness of PuT by investing in soft ICT based measures: going from words to actions under an austerity backdrop – Thessaloniki's case, Greece. Res. Transp. Econ. **51**, 40–48 (2015)
7. Meyer, M., Siwek, S., Berman, W. (Eds.) The state-of-the-practice of travel demand management. In: Transportation Research Board, TDM Innovation and Research Symposium, Transportation Research Circular 433, Washington (1994)
8. Marshall, S., Banister, D., McLellan, A.: A strategic assessment of travel trends and travel reduction strategies. Innov. Eur. J. Soc. Sci. **10**, 289–304 (1997)
9. Marshall, S., Banister, D.: Travel reduction strategies: intentions and outcomes. Transp. Res. Part A **34**, 321–338 (2000)

10. Evans, M., Davies, J.: Understanding policy transfer: a multi-level, multi-disciplinary perspective. Public Adm. **77**(2), 361–385 (1999)
11. James, O., Lodge, M.: The limitations of 'policy transfer' and 'lesson-drawing' for public policy. Res. Polit. Stud. Rev. **1**(3), 179–193 (2003)
12. Macário, R., Marques, C.F.: Transferability of sustainable urban mobility measures. Res. Transp. Econ. **22**(1), 146–156 (2008)
13. Marsden, G., Stead, D.: Policy transfer and learning in the field of transport: a review of concepts and evidence. Transp. Policy **18**(3), 492–500 (2011). ISSN 0967-070X
14. Misra, A., Gooze, A., Watkins, K., Asad, M., Le Dantec, C.A.: Crowdsourcing and its application to transportation data collection and management. In: TRB 2014 Annual Compendium of Papers. Georgia Institute of Technology (2014)
15. Slotterback, C.S.: Public involvement in transportation project planning and design. J. Archit. Plan. Res. **27**(2), 144 (2010)
16. Rose, G., Ampt, E.: Travel behavior change through individual engagement. In: Hensher, D.A., Button, K.J. (ed.) Handbook of Transport and the Environment (Handbooks in Transport), vol. 4, pp. 739–755 (2003)
17. Rudinger, G., Donaghy, K., Poppelreuter, S.: Societal trends, mobility behavior and sustainable transport in Europe and North America: the European Union network STELLA. Eur. J. Ageing **1**, 95–101 (2004). https://doi.org/10.1007/s10433-004-0009-y
18. Garling, T., Fujii, S.: Travel behavior modification: theories, methods, and programs, Chap. 5. In: The Expanding Sphere of Travel Behavior Research. Emerald Group Publishing (2009)
19. Verplanken, B., Walker, I., Davis, A., Jurasek, M.: Context change and travel mode choice: combining the habit discontinuity and self-activation hypotheses. J. Environ. Psychol. **28**(2), 121–127 (2008)
20. Gärling, T., Bamberg, S., Friman, M., Fujii, S., Richter, J.: Implementation of soft transport policy measures to reduce private car use in urban areas. In: Panels of the Energy Efficiency and Behavior Conference. European Council for an Energy Efficient Economy (2009)
21. Ceresia, F.: Interactive learning environments (ILEs) as effective tools for teaching social science. Procedia – Soc. Behav. Sci. **217**, 512–521 (2016)
22. De Gloria, A., Bellotti, F., Berta, R., Lavagnino, E.: Serious Games for education and training. Int. J. Serious Game **1**(1), 1–14 (2014). https://doi.org/10.17083/ijsg.v1i1.11
23. Martens, A., Diener, H., Malo, S.: Game-based learning with computers – learning, simulations, and games. In: Pan, Z., Cheok, A.D., Müller, W., El Rhalibi, A. (eds.) Transactions on Edutainment I. LNCS, vol. 5080, pp. 172–190. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69744-2_15
24. MOTIVATE project: D3.2.1/design of the logical architecture of MOTIVATE platform, Interreg-MED programme (2017)
25. Wagner, J.: Measuring the performance of public engagement in transportation 575 planning: three best principles. In: TRB 2013 Annual Meeting, vol. 954 (2013)
26. Dennis, P.E., Wallace, R., Reed, B.: Crowdsourcing transportation systems data. In: Michigan Department of Transportation (MDOT) REQ. NO. 1259, Connected and Automated Industry Coordination Sequence D 02 Crowd Sourced Mobile Applications (2015)
27. Dennis, E.P., Cregger, J., Hong, Q.: ITS data ethics in the public sector. Center for Automotive Research and Michigan Department of Transportation (2014)
28. Efthymiou, D., Antoniou, C.: Use of social media for transport data collection. Procedia – Soc. Behav. Sci. **48**, 775–785 (2012)
29. Vlahogianni, E., Karlaftis, M., Golias, J.: Short-term traffic forecasting: where we are and where we're going. Transp. Res. Part C: Emerg. Technol. Part **43**(1), 3–19 (2014)

30. Tomasic, A., Zimmerman, J., Garrod, C., Huang, Y., Nip, T., Steinfeld, A.: The performance of a crowdsourced transportation information system. In: TRB 2015 Annual Meeting (2015)
31. Figliozzi, M., Blanc, B.: Evaluating the use of crowdsourcing as a data collection method for bicycle performance measures and identification of facility improvement needs. Report No. FHWA-OR-RD-16-04 (2015)
32. Hood, J., Sall, E., Charlton, B.: A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. Int. J. Transp. Res. **3**, 63–75 (2011)
33. Cycle Atlanta (2013). http://www.cycleatlanta.org. Accessed July 2017
34. Jette, A., Breck, A., Johns, R.: Integrating balanced scorecard performance management with crowdsourced strategic planning. In: TRB 2015 Annual Meeting (2015)
35. Steiger, D.: Democratizing strategy: how crowdsourcing can be used for strategy dialogues" 10 California Management Review; Summer 2013, 54:4, pp. 44–69 (2013)
36. Banister, D.: The sustainable mobility paradigm. Transp. Policy **15**(2), 73–80 (2008)
37. Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C., Sengers, P.: Sustainably unpersuaded: how persuasion narrows our vision of sustainability. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 947–956. ACM (2012)
38. Froehlich, J., Dillahunt, T., Klasnja, P., Mankoff, J., Consolvo, S., Harrison, B., et al.: UbiGreen: investigating a mobile tool for tracking and supporting green transportation habits. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2009, pp. 1043–1052. ACM (2009)
39. Broll, G., Cao, H., Ebben, P., Holleis, P., Jacobs, K., Koolwaaij, J., et al.: Tripzoom: an app to improve your mobility behavior. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM 2012), (Article 57). ACM, New York (2012)
40. Gabrielli, S., Forbes, P., Jylh, A., Wells, S., Sirın, M., Hemminki, S., Nurmi, P., Maimone, R., Masthoff, J., Jacucci, G.: Design challenges in motivating change for sustainable urban mobility. Comput. Hum. Behav. **41**, 416–423 (2014)
41. Beirão, G., Cabral, S.J.: Understanding attitudes towards PuT and private cars: a qualitative study. Transp. Policy **14**(6), 478–489 (2007)
42. Edwards, W.: Dynamic decision theory and probabilistic information processings. Hum. Factors: J. Hum. Factors Ergon. Soc. **4**(2), 59–74 (1962)
43. Quadrat-Ullah, H., Karakul, M.: Decision making in interactive learning environments: towards an integrated model. J. Decis. Syst. **16**(1), 79–99 (2007)
44. Sterman, J., et al.: Management flight simulators to support climate negotiations. Environ. Model Softw. **44**, 122–135 (2013)
45. Forrester, J.W.: Industrial Dynamics. MIT Press, Cambridge (1961)
46. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for a Complex World. Irwin/McGraw-Hill, Boston (2000)

Decision Support Systems in Management and Organizations

Skills and Mindsets for an Analytically Innovative Organisation

Yusuf-Ali Karbelkar and Mike Hart(✉)

Department of Information Systems, University of Cape Town, Cape Town, South Africa
yakkoh@gmail.com, mike.hart@uct.ac.za

Abstract. Much research has been carried out on factors influencing acceptance, success, and pervasiveness of business intelligence (BI) and analytics in organisations. This study focusses on the range of skills/competencies, and attitudes/mindsets, which can lead to an organization becoming sustainably data-driven and innovative through BI and analytics (BI&A). Following a literature review, business managers, users and creators of BI&A systems in South Africa were interviewed about the contribution of skills and mindsets to pervasive BI&A. Thematic analysis of the conversations uncovered five main themes: Good Communication, Prioritization of BI, Different Skillsets, Adopting BI&A, and Hunger for BI&A, all with sub-themes. Many of the important skills and mindsets were shown to be common to BI&A practitioners, as well as to their users and management.

Keywords: Business intelligence and analytics · Skills · Mindsets Competencies · Attitudes · Innovative · Organization · Sustainable

1 Introduction

In the 4th international survey of analytics by the MIT Sloan Management Review [1] the organisations of over 2,000 business executives, managers and analysts were classified into three groups: Analytical Innovators (12%), Analytical Practitioners (54%) and Analytically Challenged (34%). Analytical Innovators have an analytics culture driven by senior mandate, are more strategic in their application of analytics, place a high value on data, and have higher levels of data management and analytic skills [2], whereas Analytical Practitioners have adequate access to data and are working to become more data driven, primarily to effect operational improvements. Their 7th survey of 2,602 respondents [2] published in February 2017 gives the split of these groups as 17%, 49% and 33% respectively. Clearly many organisations have a long way to go to become "analytically innovative".

There are other ways of classifying organisations in terms of the extent of their usage of business intelligence and analytics (BI&A), for example pervasiveness and maturity models. There are numerous definitions of business intelligence, analytics, big data etc., which change over time and according to experience of users and organisations. Here we use the term business intelligence and analytics (BI&A), as commonly understood in industry, and regularly used in major international surveys by organisations such as

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 103–118, 2018.
https://doi.org/10.1007/978-3-319-90315-6_9

Gartner. Job roles in BI&A vary from information acquisition, storage, integration, analysis, interpretation, and presentation to business analysis, management, strategy, education, training, systems implementation, change management, data science, etc. Research has shown that many different factors contribute to the level of BI&A usage and success, and that these are largely people-related and organisational rather than technological, concerning business management and users as well as BI&A practitioners [3, 4]. In this research we focus on the range of skills (or competencies) and the mindsets (or attitudes) of management and employees across an organisation, that can assist sustainable data-driven decision-making and analytical innovation. Information was gathered by interviewing knowledgeable BI&A practitioners, consultants and business management.

Structure of this paper: A brief literature review lists some of the perspectives related to this topic. The research methodology is then covered, followed by the findings from a thematic analysis of interviews with BI&A practitioners and consultants and business management. After a discussion the paper concludes.

2 Some Related Literature

For the 8th successive year, the Society for Information Management has rated BI&A (and related areas of Data Mining, Forecasting and Big Data) as the largest IT Investment in Organisations [5], and 44% of respondents currently say this investment should increase further. For the last four years the 2nd or 3rd "Personally Most Important/Worrisome IT Management Issue" of IT Leaders has been "IT Talent/Skill Shortage/Retention". After a decade of her surveys of BI adoption, Howson [6, p. 1] notes that "*Pervasive business intelligence remains elusive, with BI and analytics adoption at about 30% of all employees*" (up from about 20% ten years ago). This is one way of gauging the limited "success" of BI&A in an organization. We shall now briefly cover a few aspects relating to our research topic. (Note also that Table 3, which later lists the emergent themes from this research, also lists some references relevant to each theme and sub-theme).

2.1 Acceptance, Adoption and Diffusion of BI&A

Various theories and models have been used over the years to examine the above, such as TAM [7] and its extensions for individuals and organisations, diffusion of innovations [8], measures of success [9], and maturity models [10, 11]. These have confirmed the importance of various factors. This research however aims to look beyond initial acceptance and adoption, to the pervasive and innovative use of BI&A across an organization, concentrating on needed skills and attitudes of its individuals.

2.2 Maturity Models and Pervasiveness of BI&A

Many different maturity models of BI&A have been developed, more recent ones taking into account the newer developments of cloud, Internet of Things, big data, and its "four V's". See e.g. [10, 11]. These models typically possess a number of levels of maturity across different organizational and technical dimensions. *"The most important factors that decided the success of BI initiatives referred* [11, p. 118], *not to the technology, but to the strong belief of all users in BI, and soft competences and skills needed for BI; e.g., culture-based on facts and knowledge, trust, human resources management, managing analytical/creative people."*

Key dimensions found in [10], that were related to skills and attitudes, included (with sub-themes): Skills and experience (Leadership and management, Hiring and identifying the right people, Soft skills, Technical skills, Data science), BI&A Culture (Top management support, Promote value delivered by BI&A), and BI&A Change Management (Organisational support, Communication plan/strategy, Skills development and training).

Some BI&A maturity models have listed "pervasive" as their highest level, e.g. [11–13]. The strongest influencers of pervasiveness in BI&A were found by [3] to be organizational (rather than technological or environmental), including: Perceived value of BI&A, Executive buy-in & involvement, Strong business focus & ownership, Education & support, and Information quality, form & availability.

2.3 Skill Sets for BI&A

Skills required for traditional BI are well known, but with the advent of big data and cloud, new technologies, data sources and expectations require new skills, such as data science [14]. Some research lists entry level skills, while other skills are expected to be obtained over time. BI&A job roles and required skill sets for the big data era are examined by [15–20]. These tend to relate to BI&A professionals, as opposed to appropriate skills for users and business management.

2.4 Management and BI&A

The role of management in creating pervasive BI&A has always been heavily emphasized. Top management championship, support, and sponsorship [11, 21–23] is critical, but often proves a major challenge [24, 25]. There is a lack of managers with analytical talent and understanding for the data-driven era [26, 27], and [28] suggests ways they can become comfortable making decisions based on analytics.

2.5 BI&A Culture and Innovation

An analytic culture, generally driven from the top [25, 29], is regularly mentioned [2, 11, 22–24, 26, 28, 30]. "*Becoming a data-driven organization will involve organizational and cultural change and innovation*" [23, p. 633], but requires *"the right people, with the right skills to effect and drive the data-driven cultural change"* [23, p. 634]. Innovation is strongly related to the business impact of BI [9, p. 259], and big data has been suggested as the *"next frontier for innovation, competition, and productivity"* [26, p. 598]. Personal innovativeness and readiness for change were shown to be strong determinants of BI acceptance [4].

3 Research Questions and Methodology

The literature has shown that the attitude, actions and behaviour of management and users, and the skills of those involved with BI&A are key in establishing and maintaining an analytically innovative organization. This research decided to focus on attitudes (or mindsets) and skills (or competencies) important for all organizational members, from the CEO to lower level users, as well as BI&A practitioners. The following research questions were developed as sub-objectives:

1. What skills and mindsets should employers consider when hiring someone who would work in BI&A or who would directly manage BI&A?
2. What skills and mindsets are important for users of BI&A and for middle and senior management in developing an analytically innovative organisation?

Because the aim was to obtain a deeper understanding of these questions, the research followed an interpretive philosophy and inductive approach, using qualitative data from interviews. The timeframe was cross-sectional and the purpose both exploratory and descriptive. Permission for the research was obtained from the university's committee for research ethics, and respondents from all organisations signed consent forms, and agreed to recording of the interviews.

3.1 Semi-structured Interviews

Thirteen semi-structured interviews were conducted with management and staff from nine organisations, averaging just over an hour each. Purposive non-probability sampling was used to select organisations known to have a wide experience of BI&A, and an effort made to choose knowledgeable BI&A practitioners and business management. The latter proved most difficult to obtain, and some snowball sampling was also used. Added value came from consultants with strong experience in applying BI&A across a range of organisations. Table 1 shows the job titles and sectors of the thirteen interviewees, with the codes given to them. Interviewees were sent a one page interview prompt sheet in advance of the interview to enable them to think about the topic areas beforehand. The interview followed a conversational format, guided by the prompt sheet. Interviewees were asked some broad questions about BI&A in their own organisation, then about their opinions on BI&A in organisations generally.

3.2 Analysis Process

The recorded interviews were transcribed, and supplemented with the researcher's notes. The data was then analysed in stages, following the six step procedure for thematic analysis of [31], and using the Atlas.ti software. Codes were initially attached to each relevant phrase in the interviews; these were later iteratively refined and grouped to form major themes and sub-themes [32]. It is acknowledged that another researcher might end up with a slightly different structure of themes and sub-themes, but these provide good insight into the main comments of those interviewed.

Table 1. Job titles and sectors of interviewees with codes given

Interviewee code	Job title	Sector
A.1	Snr BI Management Consultant	Consulting (International)
A.2	Consultant	Consulting (International)
A.3	Senior Consultant	Consulting (International)
B.1	IT Manager	FMCG (International)
B.2	BI Manager	FMCG (International)
B.3	Insights Manager	FMCG (International)
C.1	BI Architect	Retail
D.1	Managing Director	Financial Services
E.1	Senior BI Manager	Retail
F.1	Senior BI Manager	Retail
G.1	Snr BI Manager & Former Consultant	Retail
H.1	CIO	Education
I.1	BI Solutions Manager	Consulting

4 Research Findings

The main themes (in bold) and sub-themes emerging from the thematic analysis are shown in Table 2. The frequencies show the number of times that a phrase relating to each sub-theme was extracted from the interviews, followed by the number of interviewees who mentioned it. The remaining frequency for each theme (in bold) counted phrases that didn't fit well into any of its sub-themes. These numbers do not indicate an order of importance, but add some additional perspective to the qualitative interview content. Because of the overlap of many of these areas, it should be noted that some sub-themes could have alternatively been placed within other themes. In addition, some of the subthemes shown could be viewed as support for, or qualifiers of, the skills and mindsets of their themes. Each theme is now discussed, in no particular order, together with some supporting quotes.

4.1 Good Communication

Unsurprisingly, all participants regarded good communication as a key skill for all managers and employees, and an important requirement when hiring someone. BI&A practitioners need to be able to listen to and gather requirements from their colleagues, managers or clients, and present results to them. The managing director of a financial services company noted:

> *"So technical brilliance can take you so far, but actually the ability to communicate we're looking for and to get good relationships with people. That's what differentiated really best BI guys from the lot of them."* (D.1).

Table 2. Themes and sub-themes that emerged from thematic analysis of interviews

Themes and subthemes	Freq	Themes and subthemes	Freq
Good Communication	**119/13**	**Adopting BI&A**	**231/13**
Appropriate Language	29/13	Change Management	51/12
Consistent Systems & Processes	20/11	Alignment	37/13
Channels for Support	65/13	Adaptability	19/5
Trust and Influence	33/8	Top Management Acceptance	118/11
Prioritisation of BI&A	**94/13**	Positive Attitude	148/13
Top Management Encouragement	147/13	Competitive Mindset	19/8
Data Responsibility	52/12	**Different Skillsets**	**50/13**
Appreciating Value of BI&A	29/8	Technical Skills	121/13
Analytic Culture	41/9	Organisational Skills	79/13
Training and Development	77/13	Analytical Skills	74/13
Hunger for BI	**91/12**	Mix of Skills	224/13
Inquisitive Mindset	118/12		
Innovative Mindset	41/11		

But even though all participants agreed on the importance of good communication for an organisation, not everyone will possess this skill.

> *"A lot of people cannot do verbal stuff. They struggle to talk. It's not always a language thing. It's sometimes just not everybody can communicate well."* (B.1).

However, through experience, one can gradually build confidence to communicate effectively by presenting work to a few people and eventually to a larger audience.

Appropriate Language. Appropriate language means communicating in terms appropriate to the recipient, depending on their level of knowledge on the matter.

> *"important how you present the information. If you don't present it the right way people won't understand and they won't feel it's relevant for them."* (B.1).

Participant B.1 also mentioned that people who do not present and summarise the information well (particularly technical employees) create a barrier to understanding, and G.1 added that the communication should be clearly presented to avoid the potential of ambiguity.

Consistent Systems and Processes. Participants A.2, B.1 and F.1 emphasised the importance of the organisation using standard BI systems and processes across the entire organisation, from C-level to operational level. This makes it easier for different roles to communicate information about the system or information received from it. Especially for a global organisation, this makes it easier to communicate any concerns, and increases trust in the system and the person communicating the information.

> *"…get the organisation aligned around that platform so that we're all speaking the same language top to bottom."* (F.1).
> *"So what is very important is that your board of directors at the head office in London, they need to use exactly the same stuff that the people on the ground in South Africa are using. … because how are you going to know what I'm talking about if you cannot see the same figures."* (B.1).

Channels for Support. Various kinds of support are needed for individuals within BI&A, within the business, and between these areas, and communication channels are vital in order to receive the support. People need skills in verbal communication, instant messaging, email, project software, etc. A.2 and G.1 emphasised importance of documenting the communicated information for future reference, and G.1 stated:

> *"Communication channels (are) very important in a project. If a technical guy needs help in something, must he go through five steps to get to the person to answer… I try to remove the obstacles of communication. It must not become a bottleneck. There mustn't be an intermediate."* (G.1).

Trust and Integrity. The trustworthiness and integrity of an individual is considered important in communication. Participant E.1 said that people would most likely communicate with an individual who is regarded as an expert on a subject, and if they trust their opinion on a matter. The added ability to solve problems increases the individual's power to influence through communication.

4.2 Prioritisation of BI&A

The prioritisation of BI&A is important to organisations, whether for operational, tactical or strategic reasons, as it will assist them to make better decisions. Organisations need to manage and take advantage of the vast amount of information now available, to be competitive and keep or increase their share of the market. (B.2).

> *"…there's a bigger priority inside most organisations to actually make sure that the BI capability is exceptional."* (C.1).

Top Management Encouragement. All participants highlighted the importance of encouragement from top management.

> *"…it is absolutely critical that your executives and probably even the CEO of an organisation buy into the BI concept, because, if that doesn't work the rest of the organisation does not follow"* (A.3).

Interviewees C.1, E.1 and D.1 stated that BI&A is used across the entire organisation and is encouraged from the top C-level to the bottom level. In contrast B.2 complained that BI&A is not taken seriously enough in their organisation, and lacks support from management to implement BI&A effectively:

> *"when I analyse data I see what's wrong with it and I know it can be fixed immediately, but it's not taken seriously and I know what value it can add and that's where I get hurt"* (B.2).

E.1 was successful in convincing higher management to “buy-in” to the idea of BI&A, and explained *“…it can be very hard to do. I’ve spent (a) long (time) working out my strategy on how to sell this as a plan*”.

Appreciating the Value of BI&A. This attitude can lead to higher prioritization of BI&A; G.1 experienced that many organisations consider BI&A as an afterthought:

> *“You get two types of systems. Those that make the wheels turn and those that watch the wheels turn. BI is watching the wheels turn. So the first systems that they put in place are to make the wheels turn”.*

F.1 stressed organisations should *“consider and make BI part of the solution before you have a solution”,* while I.1 agreed *“where we’ve had success is where we’ve been involved right from the start, and where we’ve been able to work from end to end”*. B. 2 noted the need to be seen as a *“strategic partner”,* not just as a *“service.”*

An Analytic Culture. An analytic culture prioritises BI&A, and encourages all decisions to be fact-based, which facilitates adoption of BI&A and growth of the necessary skills. Adoption of BI&A by top management can set an analytical culture through the organisation.

> *“That’s a report that lots of senior executives, C-level, they go in and they run that report every day. … if you’ve got good management structures, and have an organisational culture that says, “Look, this is how we offer it,” then people very quickly adopt those ways of working.”* (F.1).

New employees should fit into the culture of the organisation “*hard to find the right people, and they’ve got to be a culture fit and they’ve got to be a team fit*” (E.1).

The organisation does not need complex BI&A reports to create an analytic culture:

> *“…you can drive the change of behaviour without very complicated BI or analysis systems or suites in place. So if you have people making decisions written off data … that is a really good start.”* (G.1).

Data Responsibility. All stakeholders, in the business and in BI&A, should value and take responsibility for the quality of the data. Suitable governance should be used to enable appropriate wide data sharing. Interviewees largely commented on cases where this was not adequately prioritised, e.g. “*…when people don’t see information as valuable and important as money they don’t care,*” and “*It took me forever because of incorrect information.*” (B.2). Also “*…they [are] making decisions on incorrect information there’s your biggest thing for a failure on BI projects now.”* (I.1).

Training and Development. An organisation prioritising BI&A would encourage training and development of all employees in the technology and philosophy of BI&A, enhancing competitive edge and organisational profits (D.1 & G.1). Some felt that while academic institutions prepare a student with specific skills and mindsets for BI&A, a mature mindset is only achieved from working in a corporate environment, and that certain skills such as change management can only be taught on the job.

4.3 Different Skillsets

An ideal employee would possess strong skills in the organisational, managerial, technical and analytical areas, but in most cases some of these are limited.

Technical Skills. Technology aspects of BI&A are constantly changing, particularly with the recent advent of cloud, Internet of Things, "big data" and new platforms.

> *"We (are) changing very fast and you no longer have the environment where you can say we (are) going to do this for the next five years"* (C.1).

Technical skills required for BI&A include proficiency in database management, technology infrastructure, integration and management, *"where the information is captured and where it is kept."* (B.1). A.2 commented that technical employees should be able to have problem-solving skills, and to work well in a team. When discussing a BI&A technical solution with a non-technical individual, they should be able to *"reduce the complexity and make it easy for them."* (E.1).

According to C.1, the rate of technology change used to be low, but now *"everything changes every three months"*. Technology needs to adapt to frequent change in organisational requirements and technology updates, therefore technical employees would continually need to learn and adapt to skills for new areas, such as Hadoop. Interviewees C.1, B.1, D.1 and E.1 all pointed out the need to provide suitable self-service BI&A.

Organisational Skills. Organisational skills include understanding and implementing the business operations, strategies and processes from an organisational viewpoint. Any organisational need is communicated to the technical and analytical employees. BI&A needs employees with sound knowledge and appreciation of the organisation to ensure that the BI&A solutions are aligned with the organisation, and a champion or sponsor can help greatly.

> *"when you do a BI project they always say find a champion in the business that's going to sponsor you … So if you can get a sponsor, your BI projects will go well."* (G.1).

Analytical Skills. Analytical skills are key determinants of BI&A success in the organisation. They include critically viewing and analysing organisational data, together with historical and external data (such as environmental or political information), in order to gain further insights. Analysts and data scientists use statistical, machine learning, and other skills and tools to analyse the data. Interviewees A.1 and F.1 state that one needs complementary organisational skills and an inquisitive mindset in order to gain beneficial insights for the organisation.

> *"…you need somebody that actually when they look at information, don't just interpret it as a number. They can actually visualise that information in a way that it would make it easier for other people to understand it and actually get value out of it"* (C.1).

Analytical skills and mindsets should be present across the organisation as a whole, and not only found in the BI&A arena. It is particularly important for management to have at least a basic understanding of some of the analytic processes and their potential value. "*You have to create an environment where people are actually using your information*" (C.1).

Mix of Skills. Ideally all technical employees should have some organisational and analytical skills. If a technical employee is aware of what they are contributing to the organisation it would increase their interest in their work from feeling part of the "bigger picture" (F.1).

> *"BI is only successful if you actually work in with the business."* (C.1).

Participant G.1 mentioned that their technical employees should be knowledgeable in organisational and analytical areas.

> *"…even your technical people … If you ask them to fix a problem they need to analyse the system not from an information point of view but from a technical point of view. They need to be able to look at the data as well and see okay so here the figures went wrong"* (G.1).

Similarly, the organisational and analytical employees should have at least some broad knowledge of technology issues. According to all participants, it is, however, difficult to find employees with a mix of all types of skills needed for the success of BI&A. Consequently, someone with "hard to find" technical skills, but lacking organisational and analytical skills might still be employed (B.1 and E.1).

4.4 Adopting BI&A

According to all participants, BI&A should be accepted by everyone in the organisation, as everyone uses information. However, acceptance requires trust and error-free data. *"The one thing that is very important is the accuracy of your information. If people believe in your information they will want to use it."* (B.1).

Interviewees noted that the level of usage and complexity of BI&A varies across organisational levels and units, with top management typically looking at high level summaries, and those at lower levels needing more detailed operational information.

Alignment. All participants stated that, when adopting any kind of information technology or process, it needs to align with the organisation's culture and goals. If not, limited value will be added.

> *"…it's critical that every project or function is ultimately to meet business requirements and to add value to the business. …. if the intended business benefits cannot be realised then the solution means absolutely nothing…"* (A.3).

Top Management Acceptance. As with most IT-related areas, top management has a significant influence on the adoption of BI&A. Encouraging top management to use BI&A themselves is important. To convince top management to plan to use BI&A or to give support to a new project, one would need to present benefits that the organisation would receive, and be prepared to answer any questions (E.1).

Change Management. Results of a BI&A initiative often imply significant changes in organisational processes and personal responsibilities. Change management skills are therefore needed when preparing for and implementing these changes, and overcoming employee resistance. According to B.1, the change management team would need to find a way to get their "*buy-in*", explain the reason behind the implementation, the benefits that the organisation and individuals would receive, and how each employee is affected by the change.

> *"you actually need the attitude of embracing what's new, have an open enough mindset about exploring options that can actually work better for you, and as long as you can get that mindset inside a person they not an obstacle"* (C.1).

G.1 stated that, because of culture barriers, mindsets vary across different organisation levels, and good communication was essential. A.2 felt that change management was often ineffective when change was imposed by top management on lower level staff, and suggested that there should be a representative of each organisational level in the change management team.

Adaptability. Adopting BI&A requires a person to be adaptable and flexible (D.1).

> *"…it's definitely an adaptable person whose mind is not closed off to growing themselves you know, because you actually looking for somebody that's prepared to grow themselves"* (B.1).

Positive Attitude. The attitude of an employee working with BI&A affects its effectiveness, according to all participants. *"…attitude is important … you can change attitudes towards your analysis by making sure that it's accurate in the first instance and that people believe in it"* (B.1)

E.1 stated that an employee working with BI&A should have a positive attitude, and be resilient to any negative attitude of users or managers challenging their BI&A. B.1 also mentioned that the attitude of BI&A stakeholders such as users and managers is made more positive by involving them in the analysis before presenting them with the final outcome. I.1 said that a positive attitude in a BI&A employee could come from having a passion or strong interest in the technology.

Competitive Mindset. A competitive mindset encourages people to adopt BI&A.

> *"BI is probably one of the components that gives an organisation a strategic advantage over other organisations… the ability to absorb information and to make use of it, that's where organisations get their real value out of BI"* (A.1).

According to H.1, the attitude towards BI&A and its level of adoption may depend on the level of competitiveness of the organisation's market sector.

4.5 Hunger for BI&A

An analytically innovative organisation tends to have *"a hunger for BI&A"*. With this mindset, organisations want to be data-driven and to base their decisions on data. Participant E.1 describes their organisation this way, and notes it is fortunate enough to have top management encouraging the use of BI&A. This hunger for BI&A goes hand-in-hand with an inquisitive mindset, which can lead to innovation.

Inquisitive Mindset. Being inquisitive means that an individual is continuously questioning current procedures, processes and systems in the organisation, and assessing if there is a better way, and a place for innovation to occur. This could happen at all levels of the organisation.

Innovative Mindset. Participants stated that innovation came from being inquisitive, creative and intelligent. Innovation in BI&A involved creating a new way of acquiring, processing, analysing or delivering information, inside or outside of the organisation. G.1 and E.1 gave examples of how the culture and hunger for BI&A in their organisations had led to successful innovations.

5 Discussion

Interviewee information was rich, covering many different organisational circumstances. Some organisations, like C, D and E, appeared to have strong support for BI&A from the top, and fairly pervasive BI&A, whereas this was limited in B. A number were expanding into the big data area, and others were still establishing good business cases for an investment in big data. Interviewees came from different business sectors, inside and outside the BI&A area, and all had managerial and corporate experience.

Skills were in a sense "more definite" than attitudes and mindsets, which proved more difficult to identify. The overall results were supported by the literature, although some sub-themes had had limited past coverage (See Table 3 for some supporting references).

In Sect. 3 two research questions were posed:

1. *What skills and mindsets should employers consider when hiring someone who would work in BI&A or who would directly manage BI&A?*
2. *What skills and mindsets are important for users of BI&A and for middle and senior management in developing an analytically innovative organisation?*

Research Question 1. Here the research aimed to identify the broad range of skills required, not details of specific technical skills such as Hadoop, Pig and Hive, or of analytical processes or tools, e.g. machine learning and R. If organisations are embarking on big data exercises their IT and BI&A staff will certainly need to obtain a range of new skills, and these can be found in related references. This does not, however, mean that they can do without the "traditional" skills of BI&A, or that they can neglect any of its desirable attitudes or mindsets.

A suggested list of skills and mindsets is shown in Table 3 overleaf. This gives all themes and sub-themes, and some related references. The "applicability" column suggests which sets of individuals should possess and exercise these skills and attitudes. B denotes BI&A practitioners, and A denotes all people in the organisation. A bold font suggests it is a strong requirement, otherwise less so.

Table 3. Themes and sub-themes, some supporting references, and applicability*

Themes and subthemes	Some references	Applicability*
Good Communication	[10, 15, 16, 19, 20, 27]	**A**
Appropriate Language	[2, 10, 28]	**A**
Consistent Systems & Processes	[2, 28]	A
Channels for Support	[1]	M, U, **B**
Trust and Influence	[2, 25, 28]	A
Prioritisation of BI&A	[10, 25]	**M,** U, **B**
Top Management Encouragement	[1, 3, 9–11, 21–25, 29]	**M**
Data Responsibility	[1–3, 16]	**A**
Appreciating Value of BI&A	[2, 3, 21, 22, 25, 26, 28]	**A**
Analytic Culture	[1, 2, 6, 22–26, 28, 29]	**A**
Training and Development	[2, 3, 10, 16, 20, 25–28]	M, U, **B**
Hunger for BI	[1, 2]	A
Inquisitive Mindset	[23]	A
Innovative Mindset	[1, 2, 4, 9, 22, 26, 28]	A
Adopting BI&A	[9, 23, 29]	**A**
Change Management	[4, 10, 16, 23, 29]	**A**
Alignment	[16, 25]	**A**
Adaptability	[2, 16, 27, 28]	A
Top Management Acceptance	[1, 3, 9, 24, 25, 29]	**M**
Positive Attitude	[11]	**A**
Competitive Mindset	[1, 2, 22, 28]	**M,** U, B
Different Skillsets	[5, 6, 15, 27, 29]	**A**
Technical Skills	[10, 14–16, 19]	**B**
Organisational Skills	[2, 6, 10, 11, 16, 19, 28]	**M, U,** B
Analytical Skills	[1, 3, 14–20, 26, 28]	M, U, **B**
Mix of Skills	[1, 5, 9, 16, 19, 25–27, 29]	**A**

*Key: M = Top Management, U = Other Management and Users, B = BI&A, A = All
Bold font = strong requirement, otherwise partial

Research Question 2. The suggested skills and mindsets are also shown in Table 3 for top management (M), and for other management and users (U), and in many cases, all individuals (A). Again, a bold font suggests when it is a strong requirement.

Limitations. This research was carried out in organisations from the main business centres in South Africa, and so conclusions in some aspects may not apply everywhere. There could have been a wider response on skills and attitudes from senior/top management, but the gap was partly filled by observations of experienced consultants. It is also quite possible that thematic analysis carried out on the interview data by another researcher might lead to a slightly different set of themes and sub-themes. In addition, the suggested "applicabilities" of the themes and sub-themes in Table 3 will vary according to the circumstances of each organisation.

6 Conclusion

This research has shown that sustainable and pervasive data-driven BI&A depends to a large extent on an organisation's individuals possessing and applying a wide range of skills and attitudes. The list differs somewhat for BI&A practitioners, for top management and for other management and BI&A users. However, there are many important skills and mindsets that all should possess to a degree. If a culture that supports innovation and inquisitive thinking is encouraged across the organisation, and suitable accurate information and skills are available, there is great scope for productive data-driven decision-making.

While BI&A is currently gaining more attention in the workplace, the main challenge still remains for top management to recognise its potential in their own organisation, among the many competing areas that concern them, and to prioritise and encourage the use of innovation and BI&A across the organisation. Sustainability of pervasive data-driven decision-making in an organisation can often be affected by a change in the CEO or management strategy, so an established BI&A culture throughout the organisation is important.

Further research (quantitative) could be carried out with a larger sample survey, asking respondents to evaluate importance, and current attainment, of each skill and attitude on a Likert scale. Results by business sector, and job role could be compared.

Acknowledgements. This work is based on research partly supported by the South African National Research Foundation.

References

1. Kiron, D., Prentice, P.K., Ferguson, R.B.: The analytics mandate. MIT Sloan Management Review, Research report, Reprint 55480, 1–22 May 2014
2. Ransbotham, S., Kiron, D.: Analytics as a source of business innovation. MIT Sloan Management Review, 2017 Data and Analytics report, pp. 1–22, 28 February 2017
3. Bijker, M., Hart, M.: Factors influencing pervasiveness of organisational business intelligence. In: BUSTECH 2013, 3rd International Conference on Business Intelligence and Technology, pp. 21–26 (2013)
4. Grublješič, T., Jaklič, J.: Business intelligence acceptance: the prominence of organizational factors. Inf. Syst. Manag. **32**(4), 299–315 (2015)
5. Kappelman, L., McLean, E., Johnson, V., Torres, R., Nguyen, Q., Maurer, C., Snyder, M.: The 2016 SIM IT issues and trends study. MIS Q. Exec. **16**(1), 47–80 (2017)
6. Howson, C., Sallam, R.L.: Survey analysis: why BI and analytics adoption remains low and how to expand its reach. Gartner Research report G00326220, 30 June 2017. https://www.gartner.com/doc/3753469 (2017)

7. Davis, F.: Perceived usefulness, perceived ease of use and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
8. Rogers, E.M.: Diffusion of Innovations. Free Press, New York (1995)
9. Howson, C.: Successful Business Intelligence, 2nd edn. McGraw Hill Education, New York (2014)
10. Muller, L., Hart, M.: Updating business intelligence and analytics maturity models for new developments. In: Liu, S., Delibašić, B., Oderanti, F. (eds.) ICDSST 2016. LNBIP, vol. 250, pp. 137–151. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-32877-5_11
11. Olszak, C.: Toward better understanding and use of business intelligence in organizations. Inf. Syst. Manag. **33**(2), 105–123 (2016)
12. Rayner, N., et al.: Maturity Model Overview for Business Intelligence and Performance Management. Gartner Inc. Research (2008). http://www.gartner.com
13. Eckerson, W.W.: TDWI's Business Intelligence Maturity Model. TDWI, Chatsworth (2009)
14. Cao, L.: Data science: a comprehensive overview. ACM Comput. Surv. (CSUR) **50**(3), 43 (2017)
15. Cegielski, C.G., Jones-Farmer, L.A.: Knowledge, skills, and abilities for entry-level business analytics positions: a multi-method study. Decis. Sci. J. Innov. Educ. **14**(1), 91–118 (2016)
16. de Jager, T., Brown, I.: A Descriptive categorized typology of requisite skills for business intelligence professionals. In: SAICSIT 2016: Proceedings of the Annual Conference of South African Institute of Computer Scientists and Information Technologists, pp. 1–10. ACM (2016)
17. De Mauro, A., Greco, M., Grimaldi, M., Ritala, P.: Human resources for big data professions: a systematic classification of job roles and required skill sets. Inf. Process. Manag. 1–11 (2017, in press)
18. Debortoli, S., Müller, O., vom Brocke, J.: Comparing business intelligence and big data skills. Bus. Inf. Syst. Eng. **5**, 289–300 (2014)
19. Deng, X.N., Li, Y., Galliers, R.D.: Business analytics education: a latent semantic analysis of skills, knowledge and abilities required for business versus non-business graduates. In: 37th International Conference on Information Systems, Dublin, pp. 1–14 (2016)
20. Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., Turetken, O.: The current state of business intelligence in academia: the arrival of big data. Commun. Assoc. Inf. Syst. **34**(1), 1–13 (2014)
21. Kulkarni, U.R., Robles-Flores, J.A., Popovič, A.: Business intelligence capability: the effect of top management and the mediating roles of user participation and analytical decision making orientation. J. Assoc. Inf. Syst. **18**(7), 516–541 (2017)
22. Marshall, A., Mueck, S., Shockley, R.: How leading organizations use big data and analytics to innovate. Strategy Leadersh. **43**(5), 32–39 (2015)
23. Vidgen, R., Shaw, S., Grant, D.B.: Management challenges in creating value from business analytics. Eur. J. Oper. Res. **261**, 626–639 (2017)
24. Boyton, J., Ayscough, P., Kaveri, D., Chiong, R.: Suboptimal business intelligence implementations: understanding and addressing the problems. J. Syst. Inf. Technol. **17**(3), 307–320 (2015)
25. Yeoh, W., Koronios, A.: Critical success factors for business intelligence systems. J. Comput. Inf. Syst. **50**(3), 23–32 (2010)
26. Carillo, K.D.A.: Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. Bus. Process Manag. J. **23**(3), 598–622 (2017)
27. Chiang, R.H., Goes, P., Stohr, E.A.: Business intelligence and analytics education, and program development: a unique opportunity for the information systems discipline. ACM Trans. Manag. Inf. Syst. (TMIS) **3**(3), Article No. 12, 1–13 (2012)

28. Ransbotham, S., Kiron, D., Prentice, P.K.: Minding the analytics gap. MIT Sloan Management Review, 16 March 2015. http://sloanreview.mit.edu/article/minding-the-analytics-gap/
29. Davenport, T.H., Harris, J.G., Morison, R.: Build an analytical culture. In: Analytics at Work, pp. 137–147. Harvard Business School Publishing Corporation, Boston (2010)
30. Popovič, A., Hackney, R., Coelho, P.S., Jaklič, J.: Towards business intelligence systems success: effects of maturity and culture on analytical decision making. Decis. Support Syst. **54**(1), 729–739 (2012)
31. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(1), 77–101 (2006)
32. Fereday, J., Muir-Cochrane, E.: Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. Int. J. Qual. Methods **5**(1), 80–92 (2008)

Computer Supported Team Formation

Georgios Stavrou[1](✉), Panagiotis Adamidis[2], and Jason Papathanasiou[1]

[1] University of Macedonia, Thessaloniki, Greece
{gstavr,jasonp}@uom.edu.gr
[2] Alexander Technological Educational Institute of Thessaloniki, Thessaloniki, Greece
adamidis@it.teithe.gr

Abstract. Composing teams may be a time consuming and complex task. In any type of teams, the adequate selection of individuals to a team may increase the intellectual growth of the team in order to cooperate and reach the established goals. However, success of the composed team is not always guaranteed. To fill this gap, researchers develop different tools aiming to help team makers to assign team members to teams, in order to satisfy their expectations. The main goal of this paper is to present a literature review on team formation methodologies, tools, and applications that have been implemented. In this paper we present an analysis of what are teams, and its social structure. Next, a literature review and the efforts of computer supported team formation are presented and finally, the efforts of the researchers to achieve the optimal result are discussed.

Keywords: Team · Team composition · Computer supported team formation
Team formation methodologies/tools

1 Introduction

Team formation is crucial because team success depends largely on the appropriate assignment of team members to teams. Thus, it is important for an effective technique to be applied, so that optimal team composition can be ensured. The fact is that all teams are neither identical nor working by following the same procedures like size, purpose, and common characteristics. Distinctions between species, size, types, styles of the teams, etc. are found. Moreover, some variables such as communication skills, experience in teamwork, personality traits, are criteria that affect team effectiveness as well [1–3].

Researchers from different disciplines, using multiple data, attempt to develop tools, techniques and methodologies in order to facilitate the process of a successful team composition. Different goals, various teaming criteria and technologies make the task hard and complicated. Computer supported team formation offers an increasing range of tools and practices that increase the probability of generating an ideal team.

2 Team Formation

Team is the primary component of the social structure, which refers to a set of people, who are very close together, for a short or long time, in order to meet basic common

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 119–131, 2018.
https://doi.org/10.1007/978-3-319-90315-6_10

goals. Occasionally, investigators defined team, depending on the goal of their research and science. Many definitions have been developed, some of which are:

Lewin defines group as "a dynamic whole based on interdependence rather than similarity" [2]. Bales define team as "any number of persons engaged in interaction with one another in a single face-to-face meeting or series of such meetings" [3]. Sherif incorporates in his definition the concept of structure "a social unit which consists of a number of individuals who stand in "more or less" definite status and role relationships to one another and which possesses a set of values or norms of its own regulating the behavior of individual members, at least in matters of consequence to the group" [4].

Francis and Young define team as "an energetic team of people dedicated to achieving common goal/s, collaborating successfully by producing high-quality results, and by drawing pleasure from it" [5]. MacGrath emphasizes interrelations and defines team as "an aggregation of two or more people who are to some degree in dynamic interrelation with one another" [6]. Adair adds the responsibility of the team members [7], and Larson and LaFasto point to the coordination of its members' activities as a necessary feature of the team, to achieve the goal [8]. Katzenbach and Smith refer to team as a small number of people with complementary skills - skills that are dedicated to a common purpose, a set of performance objectives and an approach for which are mutually responsible [9]. Mankin et al. add to team definition the element of interdependence of people's activities, while the work of each member is dependent on the work of other members [10].

Keyton focuses on shared tasks and goals defining team as "three or more people who work together interdependently on an agreement upon an activity or goal", while Gould defines team as "individuals who stand in certain relations to each other, for example, as sharing a common purpose or having a common intentionality, or acting together, or at least having a common interest" [11].

Goal is the central component around which a team is formed. Consequently, the formation of each team differs from another in size, type, specific features, etc.

Definitions are numerous, according to each researcher's study, and the general characteristics of the team that is related to his work.

3 Social Structure Formation

Teams, according to the specific characteristics such as the purpose, composition, duration, etc. are classified into categories and described by different social structure. Basically, the social structure types are:

Teams: Teams are a set of individuals planned to collaborate together on a well-defined task. A team is task-oriented and limited by the scope of its goal. It expands from a pair to as large as needed. Its structure depends on the purpose of the task.

Specific-Task Team (or **Knot**), is a short period of time limited team convened to accomplish a critical goal (i.e. chat groups, online with video to collaborate for a specific topic, etc.). For Engeström [12]: 'The notion of knot refers to rapidly pulsating, distributed, and partially improvised orchestration of collaborative performance between otherwise loosely connected actors and activity systems. Knot working is characterized by a movement of tying, untying and retying together seemingly separate threads of activity".

Networks: There are two common sorts of networks. **Intentional Networks** (IN): informal, with low degree of cohesion, no need to be aware of each other. Network coordinator manages all communication needed. Members are task focused and collaborate to carry out a specific task [13]. **Social Networks** (SN): A social network is "a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, co-working or information exchange." [14].

Communities: Which are divided into: **Communities of Practice:** "CoP are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" [15]. Members of the CoP have meetings because they want to, whenever they want to, discussing, sharing aspirations, helping each other to solve problems, developing practices, personal relationships, etc. **Virtual Communities:** Virtual communities are: "an aggregation of individuals or business partners who interact around a shared interest, where the interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms [16], "community of people sharing common interests, ideas, and feelings over the Internet or other collaborative networks". Rheingold [17] defines virtual communities as "social aggregations that emerge from the Internet when enough people carry on public discussions long enough and with sufficient human feeling to form webs of personal relationships in cyberspace. A virtual community is a group of people who may or may not meet one another face to face, and who exchange words and ideas through the mediation of computer bulletin boards and networks'.

4 Team Member Characteristics

Team work is a necessary prerequisite for constructive completion of many projects. Successful team composition has become the point of focus of many studies that attempt to identify all the factors that are needed in order to compose the optimal functional team. However, it is generally accepted that it is not always possible to compose a successful team. The existence of many criteria and the complexity of their combination is a time consuming procedure, and frequently the creation of functional teams is not guaranteed.

Generally, to compose a successful team, it's crucial to consider many characteristics of team members, since every team has its special characteristics (goal, subject-matters, member's motivation, etc.). It is acknowledged that regardless of knowledge and expertise, many fundamental criteria of team members are necessary to be taken into consideration for team assignment, such as collaboration, communication skills, teamwork capability, experience [18], and flexibility.

Also, while interaction between team members is occurring, personal characteristics have to be considered as well. For a high functional teamwork, in the process of recruiting appropriate (suitable) members for a certain project, a combination of the following characteristics must be considered: interdependence, open communication, positive feedback, and commitment to team success, shared goals, interpersonal skills, functional expertise, interests, and personality traits [18, 19]. Researchers, usually, do not study all the characteristics at the same time, but a combination of them. In literature, communication is most often referred to, as the key component for building a successful team. Two aspects of communication barriers are highlighted: communication networks established in teams and difficulties encountered by team members in communicating with each other.

According to Blackwell [19], the main characteristics of a team, are communication skills, motivation through job assignment, expertise in individual specialty areas, personal attributes, and performance. He also argues that communication error is a central problem for team designing. Campion et al. [20], Taylor [21] refers also to communication, expertise, and flexibility in job assignments. Allen [22] cites that many studies indicate the positive impact of communication to project performance. For team designing MacDonald [23] included communication among the characteristics of leadership, coordination, cohesiveness, conflict, team age, and size.

For an effective team, mature communication was referred by Sundstrom et al. [24] among factors such as personality traits, organizational structure, culture, experience, and group stability over time. Prince et al. [25] identified communication as one of the skills of team process behaviours. The other is leadership, situation awareness, assertiveness, decision-making, mission analysis, and adaptability/flexibility.

Prasad [26] points out that an ineffective communication environment reduces the project team performance constraining the communication channels for idea exchanges across organizational lines. Smart and Barnum [27] stated that poor or inadequate communication often lead teams to failure. From the same point of view, Safoutin and Thurston [28] emphasized to the importance of communication in team designing, while communication errors, at key decision points, cause a great number of failures. Pawar and Sharifi [29] argue that team performance is improved when team members with adequate communication skills impel agreement between team members. For Lappas et al. [18] expert knowledge is considered in relation with communication, collaboration and team working skills.

Other characteristics such as team size, demographic, abilities (i.e. cognitive ability) have been also investigated by researchers with conflicting (positive, negative or none) correlation results to team effectiveness. In summary and to most researchers five important characteristics of team members, contribute to a successful team formation which are: Functional expertise, teamwork experience, communication skill, Flexibility and personality traits.

5 Computer Supported Team Formation Existing Applications

Adequate team composition meets difficulties during the design of a team, while forming successful teams is a complex task. Using many variables may create constraints that require complex correlations. Methods for team formation implement formal and human-oriented tasks. In formal tasks, efforts have been made to help team makers using computer science.

Computer supported tools are of great value for decision makers for team composition. Researchers focus on investigating techniques for automating team formation with computer supported techniques. These techniques vary according to team characteristics that are taken into consideration.

Mainly, most applications developed, concern the field of learning. Cruz and Isotani [30] overviewed team formation in collaborative Learning with computer support. Based on their algorithm or methodologies they used we classify them into three categories.

5.1 Multi-agent, Agent Based Algorithms

Opportunistic Group Formation was an idea of Ikeda et al. [31]. Opportunistic Group Formation is the procedure of forming, dynamically, a collaborative learning group. When the system detects that it's the right time for a learner to change to collaborative learning mode from individual mode, it forms a learning group. Each member of this learning group "is assigned a reasonable learning goal and a social role which are consistent with the goal for the whole group". The OGF system concentrates to form a learning group by negotiating with all learners' agents, taking into account individual goals, as well. When the system comes to an agreement, a learning group is formed.

I-MINDS: Soh et al. [32] worked on a computer-supported cooperative learning system in education. They developed an infrastructure called the Intelligent Multi-agent Infrastructure for Distributed Systems in Education. I-MINDS consist of three different intelligent agents: teacher, group, and student agent/s. A teacher agent, interacting with a teacher, is responsible for disseminating information streams to student agents, maintaining profiles for all students, assessing the progress and participation of different students, ranking and filtering of the questions asked by the students, and managing the progress of a classroom session. A student agent, on the other hand, mainly works as a personal helper to a specific student. The student agent, also, presents the learning material to the student and forms coalitions with other students for collaborative learning. A group agent forms and conducts structured cooperative learning such as the Jigsaw model [33]. The model monitors and facilitates group activities. These three agents cooperate to support student-student and student-instructor interaction in a typical classroom, in the distance education setting. Results in I-MINDS evaluation have shown that "the system can be used in a real time environment to support student cooperative activities.

Maghami and Sukthankar [34] introduced an agent-based simulation for examining the effects of stereotypes on task-oriented group formation and network evolution. They demonstrate that stereotype value judgments can have a negative impact on task performance, even in the mild case when the agents' willingness and ability to cooperate is not impaired. By modifying the social network from which groups are formed in a systematically suboptimal way, the stereotype-driven agents eliminate the skill diversity required for successful groups by driving the network toward specific topological configurations that are ill-suited for the task. Their results showed that making connections with agents solely based on group membership yields a sparser network with many isolated nodes.

5.2 Heuristic Algorithms

CATME is a peer evaluation system. It's an online free tool that helps instructors to form student teams and to evaluate individual performance within those teams (through peer evaluation) distribution, the particular system analyzes given data by students through a web interface, in order to facilitate grade adjustments for an equal among team members. Layton et al. [35] estimate that Team-Maker Comprehensive Assessment of Team-Member Effectiveness is a strong support system for team-based and cooperative learning and for a variety of research purposes and they suggest a combined use of them.

Graf and Bekele [36] suggested a mathematical approach for heterogeneous groups taking into account the students' personality traits, and their performance on the domain. They used an Ant Colony Optimization algorithm aiming to assign each student to the appropriate team, maximizing the diversity of the team, while at the same minimizing the deviation between the groups to minimum, in order to maximize the heterogeneity of teams formed.

DIANA: Suggested by Wang et al. [37] is a computer-supported heterogeneous grouping system that uses genetic algorithms to achieve fairness, equity, flexibility, and easy implementation. The system uses student's characteristics, to form heterogeneous groups, from 3 to 7 members, trying also to avoid the creation of exceptionally weak teams. Overall, DIANA results indicate that teams: (a) performed better than the randomly assigned groups and (b) showed less inter-group performance variance.

FROG: Craig et al. [38] using an evolutionary algorithm, suggested a mathematical model for homogeneous and heterogeneous groups in education aiming to optimize group forming. Group formation criteria, fitness measures and a set of attribute types were defined according to teachers' criteria. These refer to attributes of the students, such as GPA, sex, or timetable.

OmadoGenesis by Gogoulou et al. [39] addresses to homogeneous, heterogeneous, and mixed group formation. It relies on student's characteristics. The system implements three types of algorithms: Genetic algorithms for mixed group formation, Homo-A, and Hete-A for homogeneous and heterogeneous group formation accordingly. Groups may be also formed either manually by the teachers or randomly. The genetic algorithm and Homo-A have similar performance for homogeneous teams, while Hete-A yields better solutions than the genetic algorithm for heterogeneous groups. For mixed groups, the genetic algorithm gives better solution.

Wi et al. [40] used a genetic algorithm and social network measures in order to select a team manager and team members to form a new team in enterprising institutions. They presented a framework for analysing the knowledge and collaboration of the candidates for managers and team members. They tested the feasibility of the model. Their results ensured their suggestion as a quantitative and systematic method that can help enterprising institutions to assign the personnel for appropriate teams.

Spoelstra et al. [41] aiming to determine a fit-value for a team of learners for a specific project, suggested an automated team formation process model for use in Social Learning Networks contexts. Self-directed learners using the model (that contains the variables: knowledge, preferences and personality) form teams for project-based learning.

5.3 Fuzzy Algorithm

Strnad and Guid [42] proposed a fuzzy-genetic algorithm for the problem of project team formation by using previous quantitative data with several modelling enhancements like derivation of personal attributes, complex attributes models and handling unnecessary over competency. Their result makes up shortages of previous models, while providing the benefit of using intelligent team management system.

Christodoulopoulos et al. [43] proposed a web-based group formation tool, which, helps the instructor to automatically structure homogeneous groups using Fuzzy C-means and heterogeneous groups which are structured by Random Selection algorithm. The tool is based on up to three criteria. In grouping process, clustering algorithm informs the student about the probability of belonging to any group. Same information is also provided to the instructor. That facilitates him/her to adjust the formation manually. The tool was evaluated with only one criterion in a number of 18 groups. Evaluation results gave efficient satisfaction.

Torres et al. [44] proposed a fuzzy-based Multi-Agent Model for team formation based on nine roles defined by Belbin typology using the strengths and ideal responsibilities for each team member role. To better balance different working teams based on existing roles by applying a fuzzy logic approach that allows classifying the role performance of each individual into the team. The team algorithm performed by team-formation agent allowed them having teams of similar performance where each student can offer their best abilities and skills according to the role-played and demonstrate the effectiveness of integrating fuzzy-based approach with MAS to team formation.

5.4 Other Computational Techniques

Expert Finder: Vivacqua and Lieberman [45] presented, a user-interface agent, often used in recommender systems. Expert Finder offers matchmaking services generating automatically user models. The system, uses Java programming, helps a novice to find an answer to a difficult question. Expert Finder classifies novice and expert knowledge, and analyses documents developed in the course of routine work. The agent automatically match-make the novice and the expert.

MATEO (Making Adapted Teams Oriented to collaboration): is a generic system that supports adaption team forming. Adapted is the term for a team that is formed by selected individuals, according to their collaboration skills, their personal profiles and team needs [46]. The system takes into account physical context information, such as team member location, infrastructure, etc. and other resources. MATEO can adequately assign team members, based on their competencies and behaviors. Moreover, MATEO assists the person in charge of forming teams, by automating part of the team-formation process.

De Faria et al. [47] in an "Introductory Computer Programming" describes and evaluates an approach for constructing groups for collaborative learning of computer programming. Groups are formed based on students' programming styles. The style of a program is characterized by simple well known metrics, including length of identifiers, size and number of modules and numbers of indented, commented and blank lines. The experiments showed that collaborative learning was very effective for improving the programming style, particularly for students that worked in heterogeneous groups.

Tobar and Freitas [48] suggested a rule-based assignment tool, in order to reduce the time consuming procedure for teachers in group formation for learning. It is based on students' IMS learner information package (IMS LIP) specification [49]. It relies on students' characteristics. The instructor defines the rule according to his/her preferences for a team. Information is manually adjusted by the instructor, if needed.

Isotani et al. [50] suggested an ontology to be used as a framework, in order to facilitate group formation and collaborative learning design. The authors used the ontology, combining theory driven group formation with collaborative activities. An interface offers to the users the possibility to identify intended goals, roles, etc. The system recommends, automatically, team members and activities need to be performed, in order to fulfil team's goals. In order to validate system's usefulness, four instructors and twenty participants carried out the ontology. By using the ontological framework, the design of team activities was better, and had a positive effect on the performance of individuals during group learning.

Project Group Assignment System by Doyle et al. [51] describes a Web-based interactive system for creating teams for project work in industry or academia. The system allows project team members and managers to use a Web interface to enter projects and preference information into a database, and then allows the project manager to execute an algorithm that composes the teams. Teams are formed based on student preferences of projects (i.e. experience, location, time, etc.).

Pollalis and Mavrommatis [52] adopted methods from Group Technology derived from Group Technology (users and learning objects) and proposed a model for distance learning environments. Their system is based on an algorithmic approach that simultaneously selects appropriate learning objects to form a corresponding educational package for each group, by assuring optimal value of user's learning.

Ounnas et al. [53] proposed a semantic web technology system companying logic programming to form learner groups, using ontologies based on Friend of a Friend (FOAF) ontology. Specifically, they extended FOAF aiming to give semantic data about the learner's profile. In this case, every learner completes his own data at any time. The proposed framework handles the group formation process based on modelling the student's features, negotiating the group formation by the student's allocation problem as a Constraint Satisfaction Problem.

Babkin et al. [54] developed an evolvable web-based semantic platform called Info Port for research team formation, using four types of ontologies: organizational, ontology of scientific areas, historical and personal. In personal ontology they included professional knowledge of the researchers, articles, competencies and skills. They classified information about researchers considering three parameters: researcher as a person, researcher as a skillful agent and researcher as a team member.

Balmaceda et al. [55] suggested an intelligent assistant agent to form collaborative groups based on psychological styles, team roles and social relationships that may affect the team's performance. The psychological styles are those proposed by Myers-Briggs (extroversion/introversion, sensing/intuitive, thinker/sentimental, judgment/perception). The proposed WCSP approach was able to combine constraints and preferences both for individuals and groups.

Table 1. Listing CSTF tools/algorithms.

Ref. num.	Primary goal	Solution approach	Grouping criteria
[38]	Define an expressive, general model that provides more control over the criteria and attributes on that are used	Evolutionary algorithm	Numeric, one-of-bin, many-of-bin, both bin types, timetable
[40]	Selection of project managers and team members	Fuzzy/genetic algorithm	Knowledge, social network measurement
[41]	Provide team formation services to individual, self-directed learners in a social learning network	Fitness algorithm	Knowledge, personality, preferences
[42]	A new fuzzy-genetic analytical model for the problem of team formation	Fuzzy-genetic algorithm	(Fuzzy modeling of) skills
[44]	A fuzzy-based multi-agent model for team formation based on 9 roles defined by Belbin	Fuzzy, multi-agent model	9 roles defined by Belbin typology
[46]	Creation of an adaptive team. Based on personal characteristics and collaboration capabilities	Grouping algorithm	Personal characteristics, collaboration capabilities
[50]	Present an ontology that works as a framework that facilitate team formation and collaboration learning	Ontology framework based on learning theories	Individual/team goals, learning strategies/ behavior, knowledge/skill etc.
[54]	Information modeling and software design in the domain of research team formation	Semantic modeling	Research work, skills

6 Discussion and Conclusions

Even though successful team formation does not always guarantee successful research outcome, it's the cornerstone on which a successful team effort can be built. The advantages of collaborative teams are many and benefit both the research area and the individuals as well. Computer supported team formation is developed by researchers aiming to succeed the best. In this paper we presented a literature review of many approaches in the field of Computer Supported Team Formation. One of our major conclusions indicates that, few applications evaluate communication impact as a factor during the formation of the team/s, and in almost none of the existing applications communication is considered as a variable.

Furthermore, few applications indicate the importance of the factors knowledge/skills and personality in the team formation process. This may be due to the difficulty to establish one's personality and outline the knowledge or skill that is possessed. Nevertheless, in recent years, emphasis has been placed on features such as personality, knowledge, skill, trust, motivation and others, which play a major role in team performance.

In addition, most systems model a few and a fixed set of parameters, mostly 3-7, which limits the types and the diversity of teams, knowledge domain, learning goals, performance in previous teamwork, specific expertise, preferred time slots, preferred projects, performance and personality traits which are often used as team formation variables. This may be due to the fact that, by increasing the set of parameters it would hinder the selection of the most suitable person to the team and it would require a more complex algorithm/framework.

Although many systems ensure satisfaction to users through negotiation, none of them discuss the efficiency of the negotiation when all members are teamed simultaneously over and above these systems which are based on self-selecting team formation models, thus limiting the efficiency approach in forming teams.

Additionally most systems fail to assign a person/user to a team when handling limited or incomplete data and in some tools not all persons/users are assigned to team/s at the end of the formation, instead these persons/users are assigned manually to some teams by swapping other persons/users which decreases the efficiency in automated formation.

However, few solutions or work are useable, and most of them are used in traditional learning settings (classroom). Also many of the proposed approaches have been tested, but especially to validate the effectiveness of algorithm rather than evaluating the effects on the performance of the team that was formed. In addition and to the best of our knowledge, only few papers conducted peer evaluation after assigning individuals to teams, but almost none of them noted any problem/s that may have arisen from the beginning of the project till the completion.

Finally, almost none of the applications revealed a success rate on completing individuals/team tasks in time and none explain what team/s did, what knowledge resources were produced and how they were applied to obtain results. Nevertheless in the last years, the current perspectives for computer supported team formation and its relationship to intelligent systems augur the emergence of new approaches to the formation of groups which will take into account other parameters such as motivation, emotion, intuition etc.

References

1. Tarricone, P., Luca, J.: Successful teamwork: a case study. In: Quality Conversations - Conference Proceedings, pp. 640–646. 25th Annual International Conference Higher Education Research and Development Society Association (HERDSA), Perth (2002)

2. Lewin, K.: Resolving Social Conflicts, Selected Papers on Group Dynamics, Lewin, G.W. (ed.), p. 184. Harper & Row, New York (1948)
3. Bales, R.F., Strodtbeck, F.L.: Phases in group problem-solving. J. Abnorm. Soc. Psychol. **46**(4), 485–495 (1951)
4. Sherif, M., Sherif, C.W.: An Outline of Social Psychology (revised edition), p. 144. Harper & Row, New York (1956)
5. Francis, D., Young, D.: Improving Work Groups: A Practical Manual for Team Building. University Associates, San Diego (1979)
6. McGrath, J.E.: Groups: Interaction and Performance, p. 8. Prentice-Hall, Englewood Cliffs (1984)
7. Adair, J.: Effective Teambuilding. Gower, Aldershot (1986)
8. Larson & Lafasto: Teamwork: What Must Go Right, and What Can Go Wrong. Sage Publications Ltd. (1989)
9. Katzenbach, J.R., Smith, D.K.: The Wisdom of Teams: Creating the High Performance Borganization. Harvard Business School Press, Boston (1993)
10. Mankin, D., Cohen, S.G., Bikson, T.K.: Teams & Technology: Fulfilling the Promise of the New Organization. Harvard Business School Press, Cambridge (1996)
11. Keyton, J.: Communicating in Groups: Building Relationships for Effective Decision Making, 2nd edn., p. 5. McGraw-Hill, Boston (2002)
12. Engeström, Y.: From Teams to Knots: Studies of Collaboration and Learning at Work, p. 194. Cambridge University Press, New York (2008)
13. Winter, M.: Developing a Group Model for Student Software Engineering Teams, in Department of Computer Science. University of Saskatchewan, Saskatoon (2004)
14. Garton, L., Haythornthwaite, C., Wellman, B.: Studying online networks. JCMC **3**(1), 1–32 (1997)
15. Wenger, E., McDermott, R., Snyder, W.M.: Cultivating Communities of Practice: A Guide to Managing Knowledge, p. 8. Harvard Business School Press, Boston (2002)
16. Porter, C.E.: A typology of virtual communities: a multi-disciplinary foundation for future research. J. Comput.-Mediat. Commun. **10**(1) (2004)
17. Rheingold, H.: A slice of life in my virtual community. In: Harasim, L.M. (ed.) Global Networks: Computers and International Communication, pp. 57–80. MIT Press, Cambridge (1994)
18. Lappas, T., Liu, K., Terzi, E.: Finding a team of experts in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 467–476. ACM, New York (2009)
19. Blackwell, G.W.: Multidisciplinary team research. In: Chubin, D.E., Porter, A.L., Rossini, F.A., Connolly, T. (eds.) Interdisciplinary Analysis and Research, pp. 103–114. Lomond, Maryland (1986)
20. Campion, M.A., Medsker, G.J., Higgs, A.C.: Relations between work group characteristics and effectiveness: implications for designing effective work groups. Pers. Psychol. **46**, 823–850 (1993)
21. Taylor, J.B.: Building an interdisciplinary team. In: Chubin, D.E., et al. (eds.) Interdisciplinary Analysis and Research, pp. 141–154. Lomond (1986)
22. Allen, T.: Organizational structure, information technology, and R&D productivity. IEEE Trans. Eng. Manag. **EM-33**(4), 212–217 (1986)
23. MacDonald, W.R.: Characteristics of interdisciplinary research teams. In: Chubin, D.E., et al. (eds.) Interdisciplinary Analysis and Research, pp. 395–406. Lomond (1986)
24. Sundstrom, E., DeMeuse, K.P., Futrell, D.: Work teams: applications and effectiveness. Am. Psychol. **45**, 120–133 (1990)

25. Prince, C., Chidester, T.R., Bowers, C., Cannon-Bowers, J.: Aircrew coordination: achieving teamwork in the cockpit. In: Swezey, R.W., Salas, E. (eds.) Teams: Their Training and Performance, pp. 329–353. Ablex, Norwood (1992). Prince, A., Brannick, M.T., Prince, C., Salas, E.: Team process measurement and implications for training. In: Proceedings of the Human Factors and Ergonomics Society 36th Annual Meeting, vol. 2, San Francisco, CA
26. Prasad, B.: Decentralized cooperation: a distributed approach to team design in a concurrent engineering organization. Team Perform. Manag. **4**(4), 138–165 (1998)
27. Smart, K.L., Barnum, C.: Communication in cross-functional teams. IEEE Trans. Prof. Commun. 19–21 (2000)
28. Safoutin, M.J., Thurston, D.J.: A communications-based technique for interdisciplinary design team management. IEEE Trans. Eng. Manag. **40**(4), 360–372 (1993)
29. Pawar, K.S., Sharifi, S.: Physical or virtual team collocation: does it matter? Int. J. Prod. Econ. **52**, 283–290 (1997)
30. Cruz, W.M., Isotani, S.: Group formation algorithms in collaborative learning contexts: a systematic mapping of the literature. In: Baloian, N., Burstein, F., Ogata, H., Santoro, F., Zurita, G. (eds.) CRIWG 2014. LNCS, vol. 8658, pp. 199–214. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10166-8_18
31. Ikeda, M., Go, S., Mizoguchi, R.: Opportunistic group formation. In: Boulay, B.D., Mizoguchi, R. (eds.) Proceedings of AI-ED 1997 Artificial Intelligence in Education: Knowledge and Media in Learning Systems, pp. 166–174 (1997)
32. Soh, L.-K., Khandaker, N., Liu, X., Jiang, H.: A computer-supported cooperative learning system with multiagent intelligence. In: AAMAS 2006, Hokkaido, Japan (2006)
33. Clarke, J.: Pieces of the puzzle: the jigsaw method. In: Sharan, S. (ed.) Handbook of Cooperative Learning Methods. Greenwood Press, Westport (1994)
34. Maghami, M., Sukthankar, G.: An agent-based simulation for investigating the impact of stereotypes on task-oriented group formation. In: Salerno, J., Yang, S.J., Nau, D., Chai, S.-K. (eds.) SBP 2011. LNCS, vol. 6589, pp. 252–259. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19656-0_36
35. Layton, R.A., Loughry, M.L., Ohland, M.W., Ricco, G.D.: Design and validation of a web-based system for assigning members to teams using instructor-specified criteria. Adv. Eng. Educ. **2**(1), 1–28 (2010)
36. Graf, S., Bekele, R.: Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 217–226. Springer, Heidelberg (2006). https://doi.org/10.1007/11774303_22
37. Wang, D.Y., Lin, S.S.J., Sun, C.: DIANA: a computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. Comput. Hum. Behav. **23**, 1997–2010 (2007)
38. Craig, M., Horton, D., Pitt, F.: Forming reasonably optimal groups:(FROG). In: Proceedings of the 16th ACM International Conference on Supporting Group Work, pp. 141–150 (2010)
39. Gogoulou, A., Gouli, E., Boas, G., Liakou, E., Grigoriadou, M.: Forming homogeneous, heterogeneous and mixed groups of learners. In: Proceedings of the ICUM, pp. 33–40 (2007)
40. Wi, H., Oh, S., Mun, J., Jung, M.: A team formation model based on knowledge and collaboration. Expert Syst. Appl. **36**(5), 9121–9134 (2009)
41. Spoelstra, H., Van Rosmalen, P., Sloep, P.: Project team formation support for self-directed learners in social learning networks. In: Proceedings of the IADIS International Conference on Web Based Communities and Social Media (ICWBC & SM 2012), pp. 89–96 (2012)
42. Strnad, D., Guid, N.: A fuzzy-genetic decision support system for project team formation. Appl. Soft Comput. **10**(4), 1178–1187 (2010)

43. Christodoulopoulos, C.E., Papanikolaou, K.A.: A group formation tool in an e-learning context. In: Proceedings of the 19th IEEE ICTAI 2007, pp. 117–123 (2007)
44. Torres, S., Salazar, O.M., Ovalle, D.A.: A fuzzy-based multi-agent model for group formation in collaborative learning environments. In: Vittorini, P., Gennari, R., Di Mascio, T., Rodríguez, S., De la Prieta, F., Ramos, C., Azambuja Silveira, R. (eds.) MIS4TEL 2017. AISC, vol. 617, pp. 3–11. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60819-8_1
45. Vivacqua, A., Lieberman, H.: Agents to assist in finding help. In: ACM CHI 2000, pp. 65–72 (2000)
46. Arias-Baez, M.P., Pavlich-Mariscal, J.A., Carillo-Ramos, A.: Forming adapted teams oriented to collaboration: detailed design and case study. Dyna **80**, 87–95 (2013)
47. De Faria, E.S.J., Coello, J.M.A., Yamanaka, K.: Forming groups for collaborative learning in introductory computer programming courses based on students programming styles: an empirical study. In: Proceedings of the ASEE/IEEE Frontiers in Education Conference, San Diego, 28–31 October 2006, pp. 6–11 (2006)
48. Tobar, C.M., de Freitas, R.L.: A support tool for student group definition. In: Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference (2007)
49. Wilson, S., Jones, P.R.: What is IMS learner information packaging? Technical report, JISC (2002)
50. Isotani, S., Inaba, A., Ikeda, M., Mizoguchi, R.: An ontology engineering approach to the realization of theory-driven group formation. Int. J. Comput.-Support. Collab. Learn. **4**(4), 445–478 (2009)
51. Doyle, K., Kroha, S., Palchowdhury, A., Xu, W.: Learning achievement based on fuzzy rules with fuzzy 2002. In: Expert Systems with the Mid-Atlantic Student Workshop on Applications, vol. 38, pp. 4368–4381 (2008). Bai, S.M., Chen, S.M.: Programming Languages and Systems Pace 26. Evaluating students' University (MASPLAS 2002)
52. Pollalis, Y.A., Mavrommatis, G.: Using similarity measures for collaborating groups formation: a model for distance learning environments. Eur. J. Oper. Res. **193**, 626–636 (2009)
53. Ounnas, A., Davis, H.C., Millard, D. E.: Towards semantic group formation. In: The 7th IEEE International Conference on Advanced Learning Technologies, pp. 825–827, (ICALT 2007) (2007)
54. Babkin, E., Karpov, N., Kozyrev, O.: Towards creating an evolvable semantic platform for formation of research teams. In: Kobyliński, A., Sobczak, A. (eds.) BIR 2013. LNBIP, vol. 158, pp. 200–213. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40823-6_16
55. Balmaceda, J.M., Schiaffino, S.N., Pace, J.A.D.: Using constraint satisfaction to aid group formation in CSCL. Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial **17**(53), 35–45 (2014)

A Group Decision-Making Model for Supplier Selection: The Case of a Colombian Agricultural Research Company

Jenny Milena Moreno Rodriguez[1], Takanni Hannaka Abreu Kang[2](✉), Eduarda Asfora Frej[2], and Adiel Teixeira de Almeida[2]

[1] Corporación Colombiana de Investigación Agropecuária (CORPOICA), Bogotá, Colombia
[2] Center for Decision Systems and Information Development (CDSID), Federal University of Pernambuco (UFPE), Recife, Brazil
takanni.kang@gmail.com

Abstract. Decisions about supplier selection are important in the management of companies, as they directly influence their business continuity. The aim in such problem is to select the most suitable supplier from a set of potential ones, a task that involves several aspects besides cost. In this paper we build a model based on partial information in the context of MAVT (Multi-Attribute Value Theory) in order to select a satisfactory laboratory's equipment supplier for an agricultural research company. Based on a set of criteria determined by specialists and alternatives available at local market, we applied the Flexible and Interactive Tradeoff (FITradeoff) method for eliciting preferences of multiple decision makers (DMs) and assist them to reach a consensus solution. The results showed the applicability of the method for aiding real-life situations since it enables DMs to consider tradeoffs amongst criteria based on a structured elicitation process, while lowering the cognitive effort required from them.

Keywords: Group decision-making · Multicriteria analysis
Partial information · Supplier selection

1 Introduction

As demand for new and better equipment increases accompanying changes in technology, the purchasing function becomes more relevant for organizations, as well as the purchasing decisions [1]. One of the most critical issues in purchasing management is the supplier selection [2, 3], since it directly influences an organization's business continuity [4].

Decision-making in supplier selection problems is well studied in the management literature [5]. It focuses on choosing a supplier that meets all demands from buyers in a long strategic process [6]. Since such problems depend on several conflicting factors besides price, they can be treated by multicriteria decision-making/-aiding (MCDM/A) models [3, 7–9]. In real world applications, nevertheless, there is often an additional complexity: taking into account the preferences of different individuals in order to reach a collective solution [10, 11]. This group decision-making process involves DMs

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 132–141, 2018.
https://doi.org/10.1007/978-3-319-90315-6_11

with distinct characteristics such as their knowledge, skills and personality [12], and structured methodologies are needed in order to facilitate the conciliation of different viewpoints [10].

MCDM/A methods have been used in real-life applications [13]; however, albeit incomplete rather than complete information is often found in practice, this is not commonly taken into account in the decision-making processes [14]. The additive aggregation model works with weighted sums producing useful and reasonably reliable data [13]. This approach is relatively easy to implement and accommodates a compensatory rationality from the DM. An important issue regarding the application of such technique, however, is the elicitation of the criteria scaling constants parameters (weights), a task that in a group decision-making context may require time and patience from DMs [15]. The FITradeoff method proposed by de Almeida et al. [5] is an alternative to solve these issues, since it elicits intervals of feasible scaling constants values in the context of MAVT and the additive model, requiring only partial (or incomplete) information from DMs.

In this paper, we build an MCDM/A model considering multiple DMs in order to select the most suitable equipment supplier in a real world case: a Colombian agricultural research company, where managers need to purchase equipment for their research laboratories. Based on a set of criteria determined by specialists and alternatives available at local market, we applied the FITradeoff method to elicit DMs' preferences and assist them to reach a consensus solution. The main theoretical contribution of our work concerns using partial information with respect to the criteria scaling constants parameters in order to support the supplier selection group decision-making process.

The next sections in this paper are organized as follows. In Sect. 2 we introduce the basic foundations of MAVT as well as FITradeoff's flexible elicitation process. In Sect. 3 the case study is presented and an MCDM/A model is developed. Results are shown and discussed in Sects. 4 and 5. Finally, Sect. 6 draws some general conclusions.

2 Materials and Methods

2.1 Basic Concepts in MAVT

By Appling MAVT to the supplier selection problem, a value function v should associate a real number $v(\boldsymbol{x}_{s_j})$ to each supplier s_j in a set S composed by m possible suppliers, in which for each $s_j \in S$ there is a vector $\boldsymbol{x}_{s_j} = (x_{s_j1}, \ldots, x_{s_ji}, \ldots, x_{s_jn})$ of the supplier's scores on each criterion C_i, $i = 1, \ldots, n$. In MAVT two preference relations are admitted, the strict preference and the indifference relations. Consider, for instance, a pairwise comparison between suppliers s_j and s_z in S. It is assumed that in such situation the DM could react in two different ways: (i) he/she strictly prefers one supplier over the other $(s_j \succ s_z \Rightarrow s_z \not\succ s_j$; or $s_z \succ s_j \Rightarrow s_j \not\succ s_z)$; or (ii) he/she feels indifferent between them $(s_j \sim s_z \Rightarrow s_z \sim s_j)$ [16]. In case (i), we have the following relations between the strict preference and the value function $v: s_j \succ s_z \Leftrightarrow v(\boldsymbol{x}_{s_j}) > v(\boldsymbol{x}_{s_z})$; or $s_z \succ s_j \Leftrightarrow v(\boldsymbol{x}_{s_z}) > v(\boldsymbol{x}_{s_j})$; On the other hand, in case (ii) we have $s_j \sim s_z \Leftrightarrow v(\boldsymbol{x}_{s_j}) = v(\boldsymbol{x}_{s_z})$.

The additive aggregation model is the most popular one for MCDM/A, and produces useful and reasonably reliable results [13]. It is a compensatory approach that evaluates trade-offs between performances on criteria: a poor performance on one criterion can be compensated by a better performance in another one [7, 17, 18]. In this paper the value function v is given by this model, so that each point $\boldsymbol{x}_{s_j}$ in the evaluation space has an associated value $v\left(\boldsymbol{x}_{s_j}\right)$ obtained according to the additive weighted value function shown in (1) [19]:

$$v\left(\boldsymbol{x}_{s_j}\right) = \sum\nolimits_{i=1}^{n} k_i v_i\left(x_{s_j i}\right) \tag{1}$$

Where v_i is the monotonic marginal value function that remaps the consequence values $x_{s_j i}$ into a 0–1 range, and k_i is the weight (from now on referred as scaling constant) of criterion C_i, for which $\sum\nolimits_{i=1}^{n} k_i = 1$ and $k_i > 0$, $i = 1, \ldots, n$.

Although this MCDM/A technique has the advantage of being easily interpreted [13] and of accommodating the compensatory rationality from the DM, the determination of its scaling constants is a crucial issue, and some procedures have been developed in order to elicit them [5, 20]. In the following we briefly introduce some of these procedures.

2.2 Procedures for Eliciting Scaling Constants Under MAVT

According to Riabacke et al. [21], the weight assigned to a criterion is a scaling factor, which associates scores for that criterion to scores for all other criteria. Methods for eliciting criteria scaling constants are compensatory, that is, the extracted information on the weights assigned by the DM implicitly determines trade-offs between the number of units of one criterion he is willing to waive in order to increase the performance of another criterion by one unit.

The tradeoff procedure [19] is used to eliciting the scaling constants of criteria in the additive model. It has a strong axiomatic foundation [20] and accommodates a compensatory rationality of the DM. To assess the scaling constants, the DM is asked about his/her preferences between two hypothetical alternatives. Its intent is to find different alternatives (and their associated consequence vectors) that are of equal value to the DM. To obtain all the scaling constants in a problem with n criteria, $n - 1$ pairs of consequence vectors of equal value need to be found [22], thereby questions are made to the DM until all points of indifference are defined. Establishing such exact indifference points, however, is a critical concern in the procedure [20], which seems difficult to implement in real world cases because of the amount of cognitive effort required from the DM [23].

In order to facilitate the elicitation process, some procedures, such as the swing [24] abdicate a strong theoretical foundation to simplify the modeling. In order to keep the strong axiomatic foundation of the classical tradeoff procedure, while improving its implementation by facilitating the interaction with the DM in real world situations, the FITradeoff method [5] proposes the concept of flexible elicitation in order to elicit scaling constants under MAVT using partial information. This elicitation process is presented in the following subsection.

2.3 Flexible Elicitation Process

During the flexible elicitation process, the FITradeoff method shows at each interaction with the DM a pair of hypothetical alternatives with respect to which he/she is expected to give a preference relation, i.e., if he/she strictly prefers one over another, or is indifferent between them. A hypothetical alternative h is built in such a way that its consequence vector $\boldsymbol{x_h} = (x_{h1}, \ldots, x_{hn})$ expresses the worst possible performance values on all criteria but one. The worst values are defined according to a local scale, by looking at the range of consequence values of the problem's alternatives for each criterion. If the performance on criterion C_i should be maximized in order to accomplish the problem's objectives, then its worst performance corresponds to the smaller value (Min_i) in the range of consequences, while the highest value among alternatives (Max_i) will correspond to the best performance. Otherwise, the smaller and higher values in the range of consequences for the criterion correspond, respectively, to the best and worst performance values. A set H is composed by the feasible hypothetical alternatives under these conditions.

The two hypothetical alternatives $h_A, h_B \in H$ being compared at each interaction are defined according to the ranking of criteria scaling constants. One of the alternatives has an intermediate performance value on a criterion C_i in a better position on the ranking, while the other has the best performance on the next criterion on the ranking, C_{i+1}, given $k_i > k_{i+1}$. The performance of all remaining criteria for both of them is assigned with the worst performance values, as previously said.

Following MAVT, the DM's preference statements will lead to either: $(i)\, h_A \succ h_B \Leftrightarrow v(\boldsymbol{x_{h_A}}) > v(\boldsymbol{x_{h_B}})$; $(ii)\, h_B \succ h_A \Leftrightarrow v(\boldsymbol{x_{h_B}}) > v(\boldsymbol{x_{h_A}})$; or $(iii)\, h_A \sim h_B \Leftrightarrow v(\boldsymbol{x_{h_A}}) = v(\boldsymbol{x_{h_B}})$. By applying Eq. (1) to the $v(\boldsymbol{x_{h_A}})$ and $v(\boldsymbol{x_{h_B}})$ values, the preference relation given by the DM leads to a relation between the criteria scaling constants k_i and k_{i+1}: $(i)\, k_i v_i(x_{h_A i}) > k_{i+1}$; $(ii)\, k_i v_i(x_{h_A i}) < k_{i+1}$; or $(iii)\, k_i v_i(x_{h_A i}) = k_{i+1}$, since the marginal value functions v_i, $i = 1, \ldots, n$, assign the worst and best performance values respectively to 0 and 1.

With these inequalities gathered from DM's preference statements, a space of weights is obtained, i.e., a set of feasible values that the criteria scaling constants can assume. At each interaction, another inequality is obtained from a new statement given by the DM, and thus the space of weights is updated. These inequalities act as constraints of a linear programming problem model that runs at each step, searching for a solution to the MCDM/A problem [5, 25, 26]. The elicitation process finishes either when a solution is found, or when the DM is not willing to give additional preference information. The process is flexible, since the DM gives as much information as he/she is willing to; and it is interactive because the whole process is conducted based on an interaction with the DM, by means of a decision support system (DSS), which is available by request to the authors on www.fitradeoff.org.

3 Case Study

In this section, we present the case of a Colombian company composed by research agricultural laboratories located in different regions of Colombia. The company needs to acquire a certain type of equipment for its laboratories in order to guarantee availability of necessary resources to develop research projects related to agricultural area, trough optimization of economic resources, improving in maintenance and purchasing process, and keeping technological innovation. A multicriteria decision model was built in order to aid this decision-making problem, which involves preferences of multiple DMs with conflicting viewpoints.

3.1 MCDM/A Model

In order to provide an MCDM/A model applied to our problematic and case study, the model was built according to the 12 steps procedure proposed by de Almeida et al. [27] illustrated in the Fig. 1.

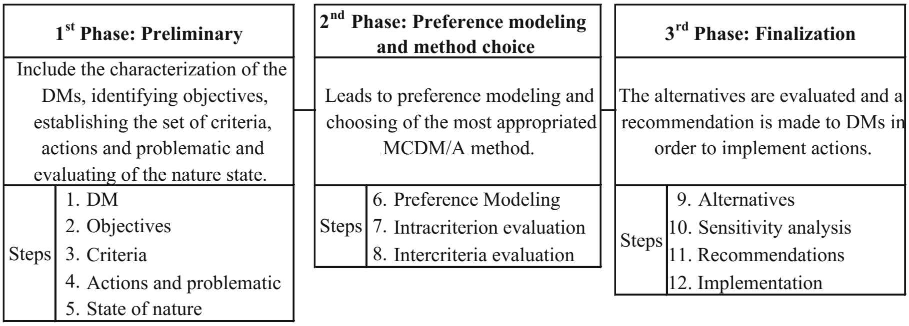


Fig. 1. Procedure for solving an MCDM/A problem. Adapted from de Almeida et al. [27].

The group decision was composed by seven people, between logistics and metrology professionals, leaders of laboratories, purchasing analysts and other equipment users from the research laboratories. Their aim is to choose a supplier that matches the best the company's needs, by taking into account different and conflicting criteria.

The criteria were selected based on judgments of the DMs, considering eight (8) of the criteria most mentioned in the literature to supplier selection, as shown in Table 1. Despite there is no standard defining the list of criteria for supplier selection [28], criteria such as quality, cost, service level, and lead time, among others, still cover the majority of those presented in the literature to be used in supplier selection [29]. In this study, criteria Price, Lead time and Technical service are to be minimized, implying that the value functions for these criteria are monotonically decreasing, while the other criteria are to be maximized.

The set of alternatives S of this problem is composed by seven potential suppliers: $s = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$, which were preselected from latest purchasing process. Finally, since we are considering deterministic variables without uncertainties, the topic associated to state of the nature does not apply in the scope of our case.

Table 1. Set of criteria considered in the problem.

Criteria	Symbol	Unit	Description
Price	C_1	$	Sale price of the equipment
Lead time	C_2	Days	Number of days between the purchase of the equipment and its installation in the laboratory
Technical capability	C_3	Units	Number of fulfilled technical specifications. The set of technical specifications comprises minimum required characteristics such as size, material, electrical system and quality norms
Warranty	C_4	Months	Number of months of responsibility of the supplier by technical failure as consequence of manufacturing or operating defects
Maintenance	C_5	Units	Number of preventive maintenance made to equipment without additional costs
Technical service	C_6	Hours	Maximum time (in hours) to answer the technical call, counted from calling to service
Payment conditions	C_7	Points	Score indicating the offered conditions by the supplier to effecting the payment
Experience	C_8	Points	Score indicating the supplier's conduct of service, considering whether the supplier had contracts and/or sales with the company before, or is able to prove the sale of same or similar equipment to other customers

Table 2. Consequences matrix.

Suppliers	Criteria							
	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
s_1	364,000	120	49	1	1	24	3	2
s_2	275,210	90	19	1	0	24	2	3
s_3	352,830	90	37	3	0	48	3	2
s_4	228,400	60	41	2	2	24	3	3
s_5	210,000	90	32	1	0	72	2	2
s_6	190,000	60	27	1	0	72	1	1
s_7	185,150	60	33	2	2	72	1	1

A compensatory rationality is considered here, in such a way that DMs allow a bad performance in one criterion be compensated by a good performance in another criterion. The MCDM/A method applied for solving this problem was the FITradeoff method [5], which has a structured elicitation process based on tradeoffs judgments in the scope of MAVT. The association of alternatives with the corresponding performance for each criterion, which is the input for the FITradeoff DSS, is shown in the consequences matrix, as showed in Table 2.

Finally, the results obtained by the FITradeoff method are presented and discussed in the next sections.

4 Results

The FITradeoff method [5] for ranking problematic was applied to each DM, in order to obtain a decreasing ranking of available suppliers for each of them. They were asked to consider tradeoffs between criteria, by comparing hypothetical alternatives and giving preference relations. The first step of the FITradeoff method is the ranking of criteria scaling constants by each DM, according to their own preferences. Table 3 shows the result of this step, with the ranking of k_i obtained by each DM at this point.

Table 3. Ranking of scaling constants k_i.

DM	Ranking
DM_1	$k_3 > k_8 > k_5 > k_4 > k_1 > k_2 > k_6 > k_7$
DM_2	$k_1 > k_3 > k_4 > k_8 > k_2 > k_6 > k_5 > k_7$
DM_3	$k_3 > k_4 > k_8 > k_6 > k_5 > k_1 > k_2 > k_7$
DM_4	$k_3 > k_8 > k_1 > k_6 > k_2 > k_5 > k_7 > k_4$
DM_5	$k_1 > k_2 > k_7 > k_8 > k_4 > k_3 > k_5 > k_6$
DM_6	$k_3 > k_4 > k_1 > k_6 > k_5 > k_8 > k_2 > k_7$
DM_7	$k_3 > k_6 > k_5 > k_4 > k_8 > k_2 > k_1 > k_7$

After this step, the DMs started to answer the elicitation questions, from which inequalities are gathered and thus the space of weights is updated at each interaction. Partial results are provided by the DSS after each interaction, so that DMs can analyze their partial rankings obtained up to that point, and also see in which position their potentially optimal alternatives are in the other DMs' rankings. This kind of analysis may shorten the elicitation process, in case the DMs are able to reach an agreement based on the visualization of partial results. Otherwise, they may continue answering questions until a complete ranking of the alternatives is found for each one of them. Table 4 the results with the final rankings obtained for each DM.

Table 4. Ranking of suppliers.

Ranking	DMs						
	DM_1	DM_2	DM_3	DM_4	DM_5	DM_6	DM_7
1st	s_4	s_7	s_4	s_4	s_7	s_4	s_1
2nd	s_1	s_4	s_3	s_1	s_6	s_3	s_4
3rd	s_3	s_6	s_1	s_5	s_4	s_7	s_3
4th	s_7	s_5	s_7	s_3	s_5	s_1	s_2
5th	s_5	s_2	s_5	s_2	s_2	s_5	s_7
6th	s_2	s_3	s_2	s_7	s_3	s_6	s_5
7th	s_6	s_1	s_6	s_6	s_1	s_2	s_6

5 Discussion of Results

From Table 4, it can be observed that supplier s_4 is in the first place of the ranking of four DMs, while s_7 is the best alternative for two DMs and s_1 is the best ranked for only one DM. Moreover, it can be seen that s_4 belongs to the second position in the ranking of two DMs, and it is in the third position for one DM. Therefore, s_4 appears to be the only alternative in the top 3 of the ranking for all DMs, and thus it seems to be a good compromise solution for this problem.

The different rankings between DMs are a result of prioritization of the various criteria, confirming that even though the criterion price is one of the most applied in practice, in the presented case study it did not have a great relevance when compared to the other criteria. Indeed, the best alternative for this problem (s_4) is not the one which offers the best sale price, indicating that other aspects besides price should be strongly considered in this kind of problem.

Although many MCDM/A methods were developed for handling real-life problems such as the supplier selection, they are not very often applied in practice, due to the difficulty for DMs to give the information required by such methods, which can be tedious and time consuming to provide [30], and may lead to inconsistencies during the decision-making process. The FITradeoff method provides to DMs an elicitation process with easier questions to answer, based on strict preference statements. Moreover, the DSS is flexible, and allows the DMs to skip difficult questions if so they desire, with another question being computed without loss of information. With an easier decision-making process, there is a reduction in the cognitive effort spent by DMs, in such a way that less inconsistency is expected in the elicitation process.

6 Conclusions

Setting a ranking between the available alternatives in purchasing under multiple criteria comprises interests and viewpoints of different groups of experts or DMs to be involved in the process. The inclusion of multiple criteria to the problem makes the decision-making process more complex, but in turns more consistent and appropriate for the problem's situation. The main goal of this study was to apply an MCDM/A approach to an equipment supplier selection problem in the context of laboratories of agricultural research, and thus determine the best alternatives by taking into account different criteria and conflicting viewpoints of multiple DMs. The additive aggregation model is suitable when the DM has a compensatory rationality and thus the FITradeoff method was applied for aiding this decision-making problem.

The elicitation process was conducted with seven DMs through a flexible process, in which partial results were available for analysis at each step, and also difficult questions could be skipped in case the DMs did not have information for answering those. A final ranking of the alternatives was obtained in the end of the elicitation process for each DM, and a final compromise solution was reached based on the analysis of the individual rankings obtained. The final solution achieved was not the one with lower sale price, which shows that a single criteria approach based only on cost factors, that often occurs in practice, may not be enough to guarantee the effectiveness of the supply chain as a whole. The FITradeoff method has shown itself to be a good tool for aiding real-life decision-making problems since it enables DMs to consider tradeoffs amongst criteria through a structured elicitation process, but at the same time lowering the cognitive effort spent by DMs.

Acknowledgments. This work was supported by the National Council for Scientific and Technological Development (CNPq) and the Coordination for the Improvement of Higher Education Personnel (CAPES). We are also grateful to Corpoica for providing the data needed to the problem and their managers for having participated in the process.

References

1. de Boer, L., Labro, E., Morlacchi, P.: A review of methods supporting supplier selection. Eur. J. Purchasing Supply Manag. **7**, 75–89 (2001). https://doi.org/10.1016/S0969-7012(00)00028-9
2. Ting, S.C., Cho, D.I.: An integrated approach for supplier selection and purchasing decisions. Supply Chain Manag.: Int. J. **13**(2), 116–127 (2008). https://doi.org/10.1108/13598540810860958
3. Amid, A., Ghodsypour, S.H., O'Brien, C.: A weighted additive fuzzy multiobjective model for the supplier selection problem under price breaks in a supply chain. Int. J. Prod. Econ. **121**(2), 323–332 (2009). https://doi.org/10.1016/j.ijpe.2007.02.04
4. Rezaei, J., Nispelinga, T., Sarkisb, J., Tavasszya, L.: A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. J. Clean. Prod. **135**, 577–588 (2016). https://doi.org/10.1016/j.jclepro.2016.06.125
5. de Almeida, A.T., de Almeida, J.A., Costa, A.P.C.S., de Almeida-Filho, A.T.: A new method for elicitation of criteria weights in additive models: flexible and interactive tradeoff. Eur. J. Oper. Res. **250**, 179–191 (2016). https://doi.org/10.1016/j.ejor.2015.08.058
6. Wetzstein, A., Hartmann, E., Benton Jr., W.C., Hohenstein, N.: A systematic assessment of supplier selection literature – state-of-the-art and future scope. Int. J. Prod. Econ. **182**, 304–323 (2016). https://doi.org/10.1016/j.ijpe.2016.06.022
7. Aissaoui, N., Haouari, M., Hassini, E.: Supplier selection and order lot sizing modeling: a review. Comput. Oper. Res. **34**(12), 3516–3540 (2007). https://doi.org/10.1016/j.cor.2006.01.016
8. Ho, W., Xu, X., Dey, P.K.: Multi-criteria decision making approaches for supplier evaluation and selection: a literature review. Eur. J. Oper. Res. **202**(1), 16–24 (2010). https://doi.org/10.1016/j.ejor.2009.05.009
9. Vinodh, S., Ramiya, R.A., Gautham, S.G.: Application of fuzzy analytic network process for supplier selection in a manufacturing organization. Expert Syst. Appl. **38**, 272–280 (2011). https://doi.org/10.1016/j.eswa.2010.06.057
10. Yoon, Y., Hastak, M., Cho, K.: Preference clustering-based mediating group decision-making (PCM-GDM) method for infrastructure asset management. Expert Syst. Appl. **83**, 206–214 (2017). https://doi.org/10.1016/j.eswa.2017.04.047
11. Zhang, Y., Xu, Z., Liao, H.: A consensus process for group decision making with probabilistic linguistic preference relations. Inf. Sci. **414**, 260–275 (2017). https://doi.org/10.1016/j.ins.2017.06.006
12. Daher, S.D.S.D., de Almeida, A.T.: The use of ranking veto concept to mitigate the compensatory effects of additive aggregation in group decisions on a water utility automation investment. Group Decis. Negot. **21**(2), 185–204 (2012). https://doi.org/10.1007/s10726-011-9266-2

13. Kaliszewski, I., Podkopaev, D.: Simple additive weighting – a metamodel for multiple criteria decision analysis methods. Expert Syst. Appl. **54**, 155–161 (2016). https://doi.org/10.1016/j.eswa.2016.01.042
14. Chen, C.T., Lin, C.T., Huang, S.F.: A fuzzy approach for supplier evaluation and selection in supply chain management. Int. J. Prod. Econ. **102**(2), 289–301 (2006). https://doi.org/10.1016/j.ijpe.2005.03.009
15. Clímaco, J.N., Dias, L.C.: An approach to support negotiation processes with imprecise information multicriteria additive models. Group Decis. Negot. **15**(2), 171–184 (2006). https://doi.org/10.1007/s10726-006-9027-9
16. Vincke, P.: Multicriteria Decision-Aid. Wiley, Chichester (1992)
17. de Boer, L., van Der Wegen, L., Telgen, J.: Outranking methods in support of supplier selection. Eur. J. Purchasing Supply Manag. **4**(2–3), 109–118 (1998). https://doi.org/10.1016/S0969-7012(97)00034-8
18. Dulmin, R., Mininno, V.: Supplier selection using a multi-criteria decision aid method. J. Purchasing Supply Manag. **9**(4), 177–187 (2003). https://doi.org/10.1016/S1478-4092(03)00032-3
19. Keeney, R.L., Raiffa, H.: Decision Making with Multiple Objectives, Preferences, and Value Tradeoffs. Wiley, New York (1976)
20. Weber, M., Borcherding, K.: Behavioral influences on weight judgments in multiattribute decision making. Eur. J. Oper. Res. **67**, 1–12 (1993). https://doi.org/10.1016/0377-2217(93)90318-H
21. Riabacke, M., Danielson, M., Ekenberg, L.: State-of-the-art prescriptive criteria weight elicitation. Adv. Decis. Sci. **2012**, 24 (2012). https://doi.org/10.1155/2012/276584
22. Keeney, R., von Winterfeldt, D.: Practical value models. In: Edwards, W., Miles, R.F., von Winterfeldt, D. (eds.) Advances in Decision Analysis: From Foundations to Applications. Cambridge University Press, Cambridge (2007)
23. Hurson, C., Siskos, Y.: A synergy of multicriteria techniques to assess additive value models. Eur. J. Oper. Res. **238**(2), 540–555 (2014). https://doi.org/10.1016/j.ejor.2014.03.047
24. Edwards, W., Barron, F.H.: SMARTS and SMARTER: improved simple methods for multiattribute utility measurement. Organ. Behav. Hum. Decis. Process. **60**, 306–325 (1994). https://doi.org/10.1006/obhd.1994.1087
25. de Almeida, A.T.: FITradeoff method for resolving evaluation of criteria by interactive flexible elicitation in group and multicriteria decision aid. In: CDSID working paper also presented as keynote at Joint International Conference of the INFORMS GDN Section and the EURO Working Group on DSS, Toulouse (2014)
26. Frej, E.A., de Almeida, A.T., Costa, A.P.C.S.: Ranking Alternatives with Flexible and Interactive Tradeoff Elicitation (2017). CDSID working paper
27. de Almeida, A.T., Cavalcante, C.A.V., Alencar, M.H., Ferreira, R.J.P., Almeida-Filho, A.T., Garcez, T.V.: Multicriteria and Multiobjective Models for Risk, Reliability and Maintenance Decision Analysis. International Series in Operations Research & Management Science, vol. 231. Springer, New York (2015). https://doi.org/10.1007/978-3-319-17969-8
28. Tadic, D., Stefanovic, M., Aleksic, M.A.: The evaluation and ranking of medical device suppliers by using fuzzy topsis methodology. J. Intell. Fuzzy Syst. **27**(4), 2091–2101 (2014). https://doi.org/10.3233/IFS-141174
29. Weber, C.A., Current, J.R., Benton, W.C.: Vendor selection criteria and methods. Eur. J. Oper. Res. **50**(1), 2–18 (1991). https://doi.org/10.1016/0377-2217(91)90033-R
30. Salo, A.A., Hämäläinen, R.P.: Preference assessment by imprecise ratio statements. Oper. Res. **40**(6), 1053–1061 (1992). https://doi.org/10.1287/opre.40.6.1053

Advances in Decision Support Systems' Technologies and Methods

SK-DSSy: How to Integrate the YouTube Platform in a Cooperative Decision Support?

Dario Torregrossa[1(✉)] and Joachim Hansen[2]

[1] Luxembourg Institute of Science and Technology (LIST), 41 Rue du Brill, 4422 Sanem, Luxembourg
dario.torregrossa@list.lu, dariotorregrossa@gmail.com

[2] Université du Luxembourg, 6 rue Richard Coudenhove-Kalergi, 1359 Luxembourg Ville, Luxembourg

Abstract. In wastewater treatment plants (WWTPs) domain, the decision support tools are nowadays necessary in order to efficiently process the large databases generated with on-line sensors. In this paper a cooperative decision support system (DSS) is presented. This DSS uses a KPI-based fuzzy logic engine to analyse the plant performance and identify the operational conditions that occur in the plants. Then, it associates the detected operational conditions with YouTube pages in which videos are uploaded to provide details and propose suggestions. The YouTube platform is then used to share and validate suggestions by means of the comment functions and the 'likes'. This approach is innovative, free of costs, and useful for plant managers that can rely on a user-friendly platform.

Keywords: Decision support system · Wastewater treatment
Fuzzy logic · YouTube
Shared Knowledge Decision Support System (SK-DSS)

1 Introduction

The energy management of Waste Water Treatment Plants(WWTPs) is generally considered a complex task because of high amount of information, complexity of phenomena, and multi-parameter optimization requirements [1]. The present paper proposes a decision support system for WWTP energy optimization that combines the fuzzy-logic analysis with the popular YouTube platform (www.youtube.com). This work is an extension of the Shared Knowledge Decision Support System (SK-DSS), the cooperative decision support system presented in [1]. Since this new tool incorporates YouTube, it has been named SK-DSSy in which 'y' stands for YouTube. Before presenting the methodology, in this introduction, the reader can find a short introduction on WWTP energy efficiency issues (Subsect. 1.1), a short description of the first SK-DSS (Subsect. 1.2), the novelty as well as the added value of this new SK-DSSy (Subsect. 1.3).

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 145–156, 2018.
https://doi.org/10.1007/978-3-319-90315-6_12

1.1 Energy Efficiency in WWTP

In the WWTP domain, several researchers have shown an interest for energy optimization because of a high energy saving potential associated with a relevant energy efficiency consumption [2]. For example, in [3], the authors investigated the efficiency of WWTPs and estimated a relevant energy saving potential (up to 25%). In [4], the authors have shown that in NW-Europe, the WWTP are mostly not energy-efficient. The magnitude of WWTP energy consumption is also relevant and, for example, in Europe, it has been estimated in 27 TWh/year [5].

1.2 SK-DSS: A Cooperative Platform for Decision Support in WWTPs

Many authors proposed methodologies to reduce the energy consumption in these facilities by optimizing the plant management. For example, Krampe proposed a benchmark analysis to identify inefficiencies [6]. Panepinto [7] used remote sensors for the on-line monitoring of electromechanical devices. In [8], the authors proposed a decision support tool for the assessment of pump performances. Poch was one of the first authors to propose an environmental decision support system to optimize the WWTP performance [9]. In line with the Poch's work, recently, the authors of the present paper have proposed SK-DSS [1], i.e. a plant generic decision support system that enables the plant operators to cooperate exchanging information. SK-DSS aims to perform analysis on a daily base using the Supervisory Control and Data Acquisition (SCADA) systems installed in the plants. The SCADA that supports SK-DSS (and its new version) is described in [5] and produces an amount of information that a human operator cannot satisfactorily use in the decision-making process [1]. SK-DSS uses a KPI-based fuzzy logic engine to analyse the plants and propose solutions. SK-DSS presented several novelties such as its plant-generic profile or the high-frequency, multi-parameter and on-line plant assessment. An extended dissertation about SK-DSS can be found in [1]; for the purposes of this paper, it is necessary to remark a specific characteristic: SK-DSS enables the operators to enlarge the list of case-based suggestions by directly upload solutions in the software databases. Then, SK-DSS automatically processes the information retrieved by the cooperative platform to provide a robust set of solutions. At the date of the first SK-DSS publication, the cooperative platform has not been fully developed yet and the operators had to upload their suggestions in an on-line PostgreSQL database. This solution was effective but not user-friendly, and some issues were still to be fixed such as the evaluation of the proposed solutions. These aspects are now improved and presented in this paper.

1.3 Novelty and Added Value of This Paper

In the present paper, SK-DSS is modified to incorporate the YouTube functions in the cooperative platform. For this reason, this evolution is called SK-DSSy. To the best of our knowledge, a decision support system that couples YouTube and fuzzy logic is new not only in the WWTP domain but also in the decision support domain in general. At the current date (12-Nov-2017), a research with key words 'YouTube' AND 'Fuzzy Logic' shows only three records in the Scopus database; these papers are not connected to decision support. No records are available in Scopus with this key-word set: 'YouTube', 'Fuzzy Logic', 'Decision Support'. Consequently, we are confident that this approach is fully original in the decision support domain.

SK-DSSy presents some added values when compared to the former version:

- it is user-friendly, because the YouTube platform is one of the mostly used over the world;
- the plant managers have not to deal with PostgreSQL to insert their suggestions;
- the knowledge is shared in a video-format; consequently video-description could be extensively adequate;
- the YouTube 'like' system can be used to give a score to the suggestions;
- the YouTube video-suggestions can be commented;
- it is possible to access the suggestions from each device connected to internet (smartphone, tablet, pc).

Moreover, SK-DSSy inherits all the characteristic of the first version: (i) it is specifically oriented to energy saving, (ii) it is plant generic, (iii) it is able to produce daily analysis reports, (iv) it provides case-based suggestions.

2 Data and Methodology

In order to provide the reader with a practical example, this paper proposes a decision support methodology to optimize the energy consumption of a blower system installed in a WWTP biological stage. This is a relevant topic because the blower energy consumption can cover up to 60% of the global energy consumption of WWTPs [11].

2.1 Data Description

The data used in this paper have been generated using the software SIMBA# (please refer to www.inctrl.ca/software/simba/). We simulated a WWTP designed for 120k population equivalent, equipped with blowers, pumps and biogas generation stage (Fig. 1). A copy of this dataset is stored in [10]. In order to show many operational conditions, the amount of air provided by the blower changes during the simulation process; according to the simulation model, these changes affect the pollution removal performance and the energy consumption.

2.2 KPI Calculation

The software is used to simulate a 200-days time-series of several parameters such as: wastewater inflow $[m^3/day]$, COD concentration at the inlet $[mg/l]$, blower

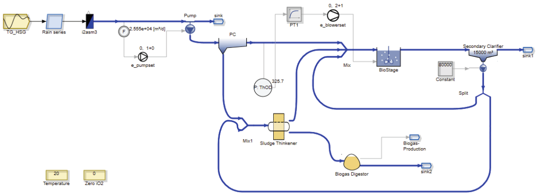


Fig. 1. Simba# simulator: screenshot

energy consumption [kWh/day], dissolved oxygen in biological stage [mg/l], COD concentration at the outlet of WWTP [mg/l]. This is the set of information required to apply the SK-DSSy methodology to blower systems. For each day and parameter, the software generates 130 values. As described in [5], these values can be aggregated in order to generate the following KPIs: population equivalent (PE), specific blower energy consumption [$kWh/year/PE$]. Moreover, as explained in [1], the SK-DSSy dynamically calculates the benchmarks taking into account the operational conditions: blower energy consumption benchmark [$kWh/year/PE$], dissolved oxygen benchmark for the activated sludge tank [fixed equal to 4 mg/l but customizable].

SK-DSS fed the fuzzy logic engine with KPIs and benchmarks, in order to be able to simultaneously work with several WWTPs [1]. Now, SK-DSSy uses benchmarks and KPIs to calculate indices that are processed by the fuzzy logic engine. In particular SK-DSSy calculates the following indices:

- BloEA_Index; equal to the ratio between the specific blower energy consumption and the blower energy consumption benchmark;
- BioDO_Index; equal to the ratio between dissolved oxygen in biological stage and the dissolved oxygen benchmark.

When the index values are higher than 1, the observed values of the parameters are higher than the benchmark. When the index values are lower than 1, the observed values of the parameter are lower than the benchmark. This changes has been done in order to improve the SK-DSSy usability because the parameters are normalized values really easy to understand.

2.3 A Plant Generic Fuzzy-Logic for Scenario Analysis

The core of SK-DSSy is the fuzzy logic engine. The reader can find an extensive mathematical explanation of fuzzy logic methodology in [12,13]. In [1,2], the reader will find two fuzzy-logic applications in WWTP domain, alongside the advantages to use this methodology in environmental domain. In particular, the fuzzy logic well performs with databases affected by uncertainty and with high-complex systems.

Table 1 shows the fuzzy rules used by SK-DSSy. These rules are obtained by processing authors' knowledge and the available literature. Each rules describes an operational scenario. For each day, the fuzzy logic engine uses these rules to analyse the indices and provide 2 outputs:

- a synthetic performance score in a range 0–100; the higher the score value, the better the plant performance;
- a truth degree (TD) for each rule in the range 0–1; the TD expresses the likelihood that in the WWTP the scenario described by the rule is verified. If, for a rule, TD = 1, then the rule describes accurately the phenomena. When TD = 0, the condition described by the rule is not verified in the WWTP.

The sum of all the truth degree for each day is equal to 1. The truth degrees of all the rules are used to identify the operational condition for given day. For example, if for one day, rule 2 [Table 1] has a TD = 0.45, rule 3 TD = 0.4, the sum of the remaining rule TD = 0.15; in this case, the energy consumption is low and the oxygen concentration is between medium and high. A potential suggestion, in this case could be to reduce a bit the air inflow to reduce energy. When the analysis is performed over a day, the interpretation of truth degree is really clear: by set-up, it is not possible that a parameter is 'low' and 'high' at the same time and consequently, in the worst case, there are 2 coherent dominant rules.

A similar investigation can be performed by analysing the average truth degree of the rules over a period. In this case, this value should be interpreted as the expected values and the logic-coherence cannot be imposed: in a period, for example, it is possible that for some days the energy consumption is low and for other days that it is high. If the analysis of a period produces two or more conflicting dominant rules, it is suggested to reduce the interval in order to have a more detailed assessment of operational conditions.

Table 1. Fuzzy logic rules

Rule	BloEA_Index	BioDO_Index	Score
1	Low	Low	30
2	Low	Medium	100
3	Low	High	80
4	Medium	Low	30
5	Medium	Medium	50
6	Medium	High	70
7	High	Low	0
8	High	Medium	20
9	High	High	40

The first column identifies the rule number. The columns 2, 3 and 4 identify the inputs. The last column identify score associated with the scenario. This table has to be translated in fuzzy logic language; for example, the rule 1 becomes:
IF BloEA_Index IS Low AND BioDO_Index IS Low THEN Score is 30

In order to use the rules, SK-DSSy needs to attribute the appropriate adjective to each parameter. For example, SK-DSSy needs to know if (or better 'how much') it is true that BloEA_Index is Low. This operation is called 'fuzzification' [12,13] and it is performed with fuzzification functions (Fig. 2).

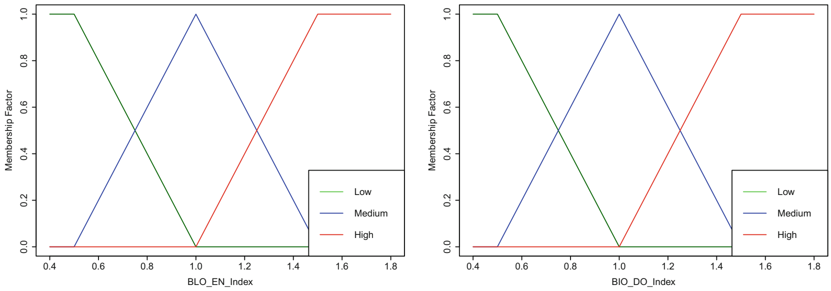


Fig. 2. Input membership functions

2.4 From Scenario Analysis to Decision Support Using the YouTube Platform

At this point, SK-DSSy explained the operational conditions occurring in the plant. This result can be coupled with a list of potential case-based solutions. In SK-DSS [1], this was done with a PostgreSQL table that plant managers can use to visualize the results and upload new solutions.

The main limitation of this approach was that plant managers should be able to use PostgreSQL queries.

A not-addressed question concerned the evaluation of the solutions proposed by the common platform; according to the philosophy of this cooperative platform, each plant operator is authorized to upload solutions. This open-access approach does not guarantee the quality of proposed solutions. The quality of the end-user contribution could be decreased by several issues, for example: a limited comprehension of the variables, the upload of plant-specific solutions or the limited experience of the end users.

SK-DSSy proposes to overcome these issues by incorporating YouTube in the cooperative platform. In SK-DSSy, the fuzzy logic rules are connected to YouTube web-pages in which the solutions can be uploaded as video or as comment. SK-DSSy identifies the dominant scenario and lead the end-user to an associated video-page. For each rule, the main video is inserted by the page administrator; it explains the operational scenario connected to the rule, some basic solutions and, in order be time efficient, it lasts less than 1 min. The end-users can visualize the videos, comment it with a text, add videos or link to external resources (such as papers or other web-pages). Another important YouTube function is the 'like' command. Each user can mark with a 'like' the useful suggestions and with a 'dislike' the comments considered not useful. The comments are consequently scored and it is possible to sort them by popularity.

The advantages of this approach are: (i) the knowledge can be shared in many formats (such as video, text or links), (ii) the use of YouTube platform is extremely user-friendly, (iii) the server and the maintenance of a part of SK-DSSy are externalized to Google-YouTube services without costs, (iv) possible assessment of the proposed solutions.

3 Results

Table 2 reports a statistical summary of the parameters connected to blower energy consumption and the dissolved oxygen concentration in the activated sludge tank. In the period under investigation (200 days), the energy consumption of the blowers as well the dissolved oxygen values vary in a large range and comprise desirable and undesirable conditions.

Table 2. Summary of KPIs

	Spec. Blower En. cons. [kWh/year/PE]	Blower En. Index [-]	DO [mg/l]
Min	0.99	0.06	0.02
1st quartile	7.51	0.49	2.76
Median	9.40	0.61	3.80
Mean	9.39	0.61	3.83
3rd quartile	12.31	0.80	5.99
Maximum	17.98	1.17	6.17

The fuzzy logic engine, for each day, produces an overall performance score (Fig. 3). For values higher than 80/100, the plant performance can be considered good, while a score value below 40/100 shows low performances of the WWTPs.

According to the SK-DSS methodology [1], it is possible to analyse single days or periods. In Fig. 4, the analysis of 4 period is provided. The rules are those of Table 1. During the first 50 days, the global index is quite high because the rule 2 is largely dominant. In the period between day 50 and day 100, the global index is in a normal range. Rules 2 and 6 show that the energy consumption is in a normal range, but it is still possible to save energy because at several day the oxygen concentration in the bio-stage can be reduced. In the third period, between day 100 and day 150, the global score is critical because we have simulated a failure of blower system. In this case, a wrong set-up of the system is detected by Rule1 that explains the bio-stage oxygen concentration is not sufficient for several days. In the last period, the global evaluation increases because rules 2, 3 and 6 have a consistent truth degree. When more rules are dominant in the system, additional analyses are required. In [2], we have shown how it is possible to produce additional analysis by investigating inflow-performance patterns. In this case, this analysis is omitted because the dataset is produced with a simulation.

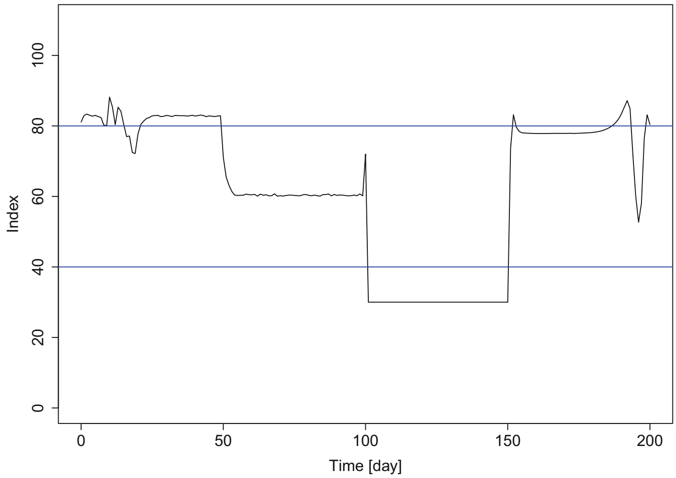


Fig. 3. Timeseries of performance score

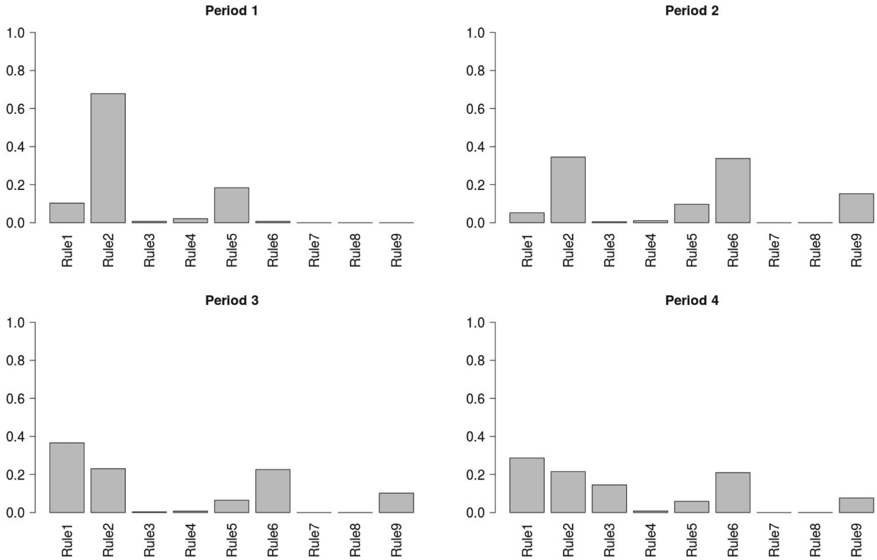


Fig. 4. Scenario analysis for many periods; Period 1, between day 0 and day 50; Period 2, between day 50 and day 100; Period 3, between day 100 and day 150; Period 4, between day 150 and day 200

3.1 From Scenario to Solutions

When the plant operators are aware of the performance of their facilities, they have to decide if and how to react. SK-DSSy supports the decision-making process by associating a video-based solution store to the detected scenarios. Table 3 shows the YouTube links associated to each rule; the reader can use these links to access the suggested solutions and interact with other end-users.

Table 3. Link to solution. The reader use the links to access the videos and interact with other end-users.

Rule	Link	Rule	Link
Rule1	Link to rule store 1	Rule6	Link to rule store 6
Rule2	Link to rule store 2	Rule7	Link to rule store 7
Rule3	Link to rule store 3	Rule8	Link to rule store 8
Rule4	Link to rule store 4	Rule9	Link to rule store 9
Rule5	Link to rule store 5		

Using these links, the end-user is redirected to the YouTube page with:

- A video that explains the operational condition of the blower system;
- The discussion between plant operators;
- Additional resources (link to papers, repository, software);
- The evaluation of each suggestion by means of the 'like' system.

Figure 5 shows a screenshot of the video connected to the rule 1. In this videos, scenario explanations with potential causes are provided.

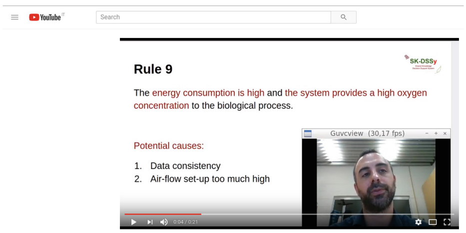


Fig. 5. Screen-shot of intro-video

In the same web-page, the comment thread is activated. Figure 6 shows for example a piece of conversation in which a sensor failure question is addressed. The comment thread can be also used to share other documents; for example, in the answer shown by Fig. 6, a link to an external book is provided. The 'like' system is used to score the answer and the contributions: the contributions with more 'like' can be sorted and visualized on the top of the page.

In summary, with SK-DSSy, the plant operators have a decision-support system able to perform the on-line analysis of WWTPs and to share knowledge. The contributions of the network can be voted and ranked. Moreover, the sharing-knowledge platform relies on YouTube platform that is well-known and generally considered user-friendly.

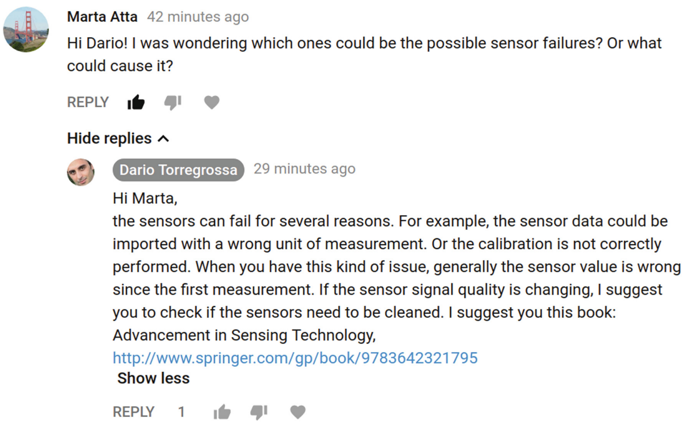


Fig. 6. Screen-shot of cooperative use of YouTube platform for decision support

4 Discussion

This paper presents an extension of the decision support system presented in [1]. The decision tools are important for WWTP operators that are required to deal with large databases (almost 300k records/day). The cooperative decision support tools enable the plant operators to cooperate and consequently the set of proposed solution is robust, self-updating and peer-reviewed. To the best of our knowledge, in water domain, SK-DSSy represents a novelty.

This methodology has been applied in WWTP domain, but it is suitable to be applied in each domain in which it is necessary to analyse a large amount of data for decision support.

This methodology has several advantages:

- it enables the on-line analysis of complex systems; in this case, a demonstration has be performed on wastewater treatment plants;
- it provides a robust and self-updating set of solutions by enabling the cooperation between experts;
- it uses YouTube, i.e. an user-friendly, cost-free, popular, and flexible platform;
- the maintenance of this cooperative platform is externalized to YouTube service providers;
- the fuzzy logic input is expressed as an index set in order to make the comprehension really easy for the end user.

Despite the advantages exposed before, the use of YouTube platform has a big limitations: it is not possible to know and/or influence the evolution of the platform during the next years. Nevertheless, given the popularity of YouTube, we expect that the service will be active at least for several years.

Moreover, a potential failure of the decision support system can be due to the proposition of wrong suggestions that are not evaluated by the end-user community. For this reason, it is necessary that end-users are stimulated to the peer-review; in a 'open' decision support system, a failure of the peer-review validation can dramatically decrease the quality of the outputs.

Another potential limitation is the absence of a score to motivate the end-users to cooperate. In the scientific domain, the impact factor is considered important for the profile of authors and it is also a stimulus to improve the quality and the quantity of contributions. A similar mechanism should be implemented to stimulate the SK-DSSy end users to cooperate.

These limitations will be overcome with a new YouTube like cooperative platform to be developed in the next years. Currently, the prototype of the SK-DSSy web-interface is available at https://dario-torregrossa.shinyapps.io/Ver2/.

In WWTP domain, the professional networks, consortia and syndicates are common: for example, Croon (NL) and Vereniging van Zuiveringsbeheerders (NL) and Wupperverband(GER); these organizations are active in collecting data and monitoring many WWTPs in The Netherlands and in Germany. SK-DSSy could be used to enhance the collaboration between plant managers inside these networks that could develop mechanisms for cooperation.

5 Conclusion

In the wastewater treatment plant domain, decision support tools are required to efficiently manage the large amount of data produced by on-line sensors. In this paper, a YouTube based cooperative decision support system has been presented. This cooperative decision support system combines the fuzzy logic analysis with the use of YouTube platform; this approach enables an efficient cooperation between plant operators and the evaluation of the proposed suggestions. Moreover, a research on Scopus database shows that this approach is extremely innovative in the water as well as in the decision support domain.

Acknowledgements. The authors gratefully acknowledge the financial support of the National Research Fund (FNR) in Luxembourg (7871388-EdWARDS) and additional funding from the Luxembourg Institute of Science and Technology. The ideas, the methodology, the script and the results presented in this paper are part of the Ph.D. project of Dario Torregrossa (EdWARDs). The authors thank Marta Antonelli (LIST) for her support in testing the platform and Ulrich Leopold (LIST) for the exchange of ideas.

References

1. Torregrossa, D., Hernández-Sancho, F., Hansen, J., Cornelissen, A., Popov, T., Schutz, G.: Energy saving in wastewater treatment plants: a plant-generic cooperative decision support system. J. Clean. Prod. **167** (2017). https://doi.org/10.1016/j.jclepro.2017.08.181
2. Torregrossa, D., Hansen, J., Hernández-Sancho, F., Cornelissen, A., Schutz, G., Leopold, U.: A data-driven methodology to support pump performance analysis and energy efficiency optimization in waste water treatment plants. Appl. Energy **111** (2017). https://doi.org/10.1016/j.apenergy.2017.09.012

3. Castellet, L., Molinos-Senante, M.: Efficiency assessment of wastewater treatment plants: a data envelopment analysis approach integrating technical, economic, and environmental issues. J. Environ. Manag. **167**, 160–166 (2016). https://doi.org/10.1016/j.jenvman.2015.11.037
4. Becker, M., Hansen, J.: Is the energy-independency already state-of-art at NW-European wastewater treatment plants. In: Proceedings of IWA-Conference Asset Management for Enhancing Energy Efficiency in Water and Wastewater Systems, pp. 19–26 (2013)
5. Torregrossa, D., Schutz, G., Cornelissen, A., Hernández-Sancho, F., Hansen, J.: Energy saving in WWTP: daily benchmarking under uncertainty and data availability limitations. Environ. Res. **148**, 330–337 (2016). https://doi.org/10.1016/j.envres.2016.04.010
6. Krampe, J.: Energy benchmarking of South Australian WWTPs. Water Sci. Technol. **67**, 2059–2066 (2013). https://doi.org/10.2166/wst.2013.090
7. Panepinto, D., Fiore, S., Zappone, M., Genon, G., Meucci, L.: Evaluation of the energy efficiency of a large wastewater treatment plant in Italy. Appl. Energy **161**, 404–411 (2016). https://doi.org/10.1016/j.apenergy.2015.10.027
8. Torregrossa, D., Leopold, U., Hernndez-Sancho, F., Hansen, J., Cornelissen, A., Schutz, G.: A tool for energy management and cost assessment of pumps in waste water treatment plants. In: Linden, I., Liu, S., Colot, C. (eds.) ICDSST 2017. LNBIP, vol. 282, pp. 148–161. Springer International Publishing, Heidelberg (2017). https://doi.org/10.1007/978-3-319-57487-5_11
9. Poch, M., Comas, J., Rodríguez-Roda, I., Sánchez-Marré, M., Corts, U.: Designing and building real environmental decision support systems. Environ. Model Softw. **19**, 857–873 (2004). https://doi.org/10.1016/j.envsoft.2003.03.007
10. Torregrossa, D.: WWTP simulation (2017). https://doi.org/10.13140/RG.2.2.24341.42727, https://www.researchgate.net/publication/320979782_Simulation_of_WWTP
11. Shi, C.Y.: Mass Flow and Energy Efficiency of Municipal Wastewater Treatment Plants. IWA Publishing, London (2011)
12. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965). https://doi.org/10.1016/S0019-9958(65)90241-X
13. Starczewski, J.T.: Advanced Concepts in Fuzzy Logic and Systems with Membership Uncertainty, vol. 284. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-29520-1

A DSS Model for Selection of Computer on Module Based on PROMETHEE and DEX Methods

Gjorgji Nusev[1,2], Pavle Boškoski[1], Marko Bohanec[1], and Biljana Mileva Boshkoska[1,3(✉)]

[1] Jožef Stefan Institute, Jamova cesta 39, Ljubljana, Slovenia
{gjorgji.nusev,pavle.boskoski,marko.bohanec,biljana.mileva}@ijs.si
[2] Jožef Stefan International Postgraduate School, Jamova cesta 39, Ljubljana, Slovenia
[3] Faculty of Information Studies in Novo mesto, Ljubljanska cesta 32a, Novo Mesto, Slovenia
http://ijs.si, http://mps.si, http://fis.unm.si

Abstract. The onset of Industry 4.0 requires for development of self-aware systems. The core of such systems are computationally powerful yet energy efficient embedded modules, called computer on module (COM), capable of performing various tasks of control, data acquisition and signal processing. The market is flooded with various COM systems thus making the selection of the most appropriate one for appropriate task is a very difficult task. This decision problem is addressed by employing PROMETHEE and DEXi methods. The proposed solution is a decision model based on 13 attributes. The evaluation is performed on a set of 53 currently available COMs.

Keywords: Decision support system · DEX · PROMETHEE
Computer on module

1 Introduction

Currently the Internet of Things (IoT) devices are covering almost every aspect of our daily life, medicine, technology, etc. This trend is becoming even more apparent with the onset of the Industry 4.0 strategy. It is a paradigm shifting process that profoundly alters the way how modern production lines operate.

It is now required to have production lines that are adaptable, self-aware of changes in their condition and fully integrated with the available information systems. Due to the harsh operating environment those information systems has to be highly specialised, endurable and robust. In addition to that the they have to be compatible with the most of the sensors available on the market today as well they need to be computationally powerful yet energy efficient embedded computer systems known as COMs or system on module. For the process of integration of a COMs into an existing production line, the following

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 157–168, 2018.
https://doi.org/10.1007/978-3-319-90315-6_13

requirements (which can also be treated as general attributes) has to be taken into an account *Security*, *Ease of development*, *Data acquisition*, *processing and storage*, *Connectivity*, *Power*, *Physical device design* and *Cost* [1].

There are a many different types of COMs available on the market today majority of which are mainly used for good for prototyping [1–3]. The problem of selecting the most appropriate COM is overwhelming for the end-user, predominantly due to the sheer number of attributes involved for describing a COM [3]. Therefore, a decision support system (DSS) designed for selecting COMs is of a significant practical merit.

Designing a DSS for industrial customers can be rather challenging. The target customers are particularly conservative and aim only for products and services that have been already thoroughly tested [4]. Furthermore, one of the most sought property is reliability and maintenance support from the supplier side [5]. Therefore, the issue of determining the most appropriate supplier is a challenging one in particular for new and emerging technology where a proper track record is missing [6,7].

The issue of selecting the most suitable COM is usually done based on cost, power consumption, performances and volume constraints [8,9]. The selection of appropriate COMs for IoT systems is particularly challenging [9]. These days the core of the most COM available are predominantly ARM based CPUs that have proved to be reliable platforms. Therefore, the main focus is on peripheral devices, connectivity, storage capacity and most importantly price and support periods of the available COMs.

Our problem involved selecting a low-cost Computer on Module with an adequate performances for our application. In the process of selection of the best COMs a database of 53 different COMs was created, which were ranked according to 13 different attributes. Due to the large number of attributes that are used for selection of the most appropriate COM, the problem can be modelled using multi-attribute decision methods.

Currently there are many different varieties of Multi Criteria Decision Methods, that have different advantages and disadvantages over the others. We looked for methods that consider both qualitative and quantitative data, require little additional information from the decision maker, methods that handle inconsistencies when the decision maker is not fully aware of his preferences and, which are freely available. The evaluation of the DSSs model was done using the Visual PROMETHEE and DEXi software, which includes the PROMETHEE and DEX method respectively [10,11]. The selection of the pre-mentioned methods was due to the fact that we considered them as the most suitable for the purpose of ranking and choosing the best COMs.

2 Description of Attributes

COMs can be characterize in the terms of device capabilities and application requirements, such as *Data acquisition and control*, *Data processing and storage*, *Connectivity*, *Power management* and *Price* [1,9]. Following the pre mentioned characterisations of COMs we build the suitable models for selection of the COMs using the following attributes:

- Random Access Memory (RAM)
- Processor speed
- Number of cores
- Flash
- Number of integrated ADC
- Number of inter-IC sound (I^2S) ports
- Number of serial peripheral interface bus (SPI) ports
- Number of universal asynchronous receiver-transmitter (UART) ports
- Number of Ethernet ports
- Number of secure digital (SD) cards
- YOCTO[1] support
- Production period
- Price

RAM is important for the process of saving the acquired data. *Flash* is used for permanently saving the acquired data, also it is a device where the operational system and other applications are saved. *Clock* represents the work frequency of the processor. The higher the frequency, the faster the computer. The number of SPI and UART communication port are required for connecting external hardware (Hard-drives, external ADC, other sensors etc.). *I2S* is an electrical serial bus interface standard used for connecting digital audio devices together, such as special microphones for vibration measurements [12]. *Ethernet* connection is required for communication via network. *SD cards* are used as an external storage.

The importance of the *support period* of the COMs used for industrial application data acquisition and analysis systems is of a great value due to its reliability. High resolution and highly accurate *ADC* (Analogue Digital Converter) is very important for the process of digitalization of the measured analogue signals, which is later used for further analysis.

In both developed models, using PROMETHEE and DEX methods, all attributes are maximized, except of the *Price*, which is minimised. This mean we are looking for best COM performances for the lowest possible cost.

When modeling with DEX, attribute values are ordered from least preferred (red) to most preferred (green) as shown in Table 1. When modeling with PROMETHEE, the attributes were grouped into the following five categories:

- CPU (RAM, Flash, Number of Cores, Clock)
- Communication (Number of SPI, I2S, Ethernet, UART, SD cards)
- Support (Support period, YOCTO support)
- Price
- Number of ADC

[1] The Yocto Project is a Linux Foundation Collaborative open source Project whose goal is to produce tools and processes that enable the creation of Linux distributions for embedded software that are independent of the underlying architecture of the embedded hardware.

Table 1. Attributes in the DEX model

№ internal ADCs	№ I^2S ports	№ SPI ports
0 0 ADC	0 0 I2S	<2 <2 SPI
1 1 ADC	1 1 I2S	2 2 SPI
>1 >1 ADC	>1 >1 I2S	>2 >2 SPI
№ ethernet ports	**Production Support**	**№ UART ports**
0 0 No ethernet	short <= 2020	<2 <2 UART
1 1 Ethernet	medium 2021 - 2025	2 2 UART
2 2 Ethernet	long > 2025	>2 >2 UART
Flash memory	**RAM**	**№ cores**
5 <128 MB	1 <128 MB	1 1 Core
4 128 MB	2 128 MB	2 2 Cores
3 256 MB	3 256 MB	4 4 Cores
2 512 MB	4 512 MB	
1 >=1024 MB	5 >=1024 MB	
Clock speed	**№ SD cards**	**Price**
1 0 - 199 MHz	1 1 SD Card	very low 0 - 50 €
2 200 - 399 MHz	2 2 SD Cards	low 51 - 100 €
3 400 - 599 MHz	3 3 SD Cards	medium 101 - 150 €
4 600 - 799 MHz	4 4 SD Cards	high 151 - 200 €
5 800 - 999 MHz		very high >201 €
6 1000 - 1199 MHz		
7 1200 - 1400 MHz		
YOCTO no/yes		

3 Modelling with PROMETHEE

The PROMETHEE and GAIA methods are among the most widely used multicriteria decision aid methods. PROMETHEE stands for **P**reference **R**anking **O**rganization **METH**od for the **E**nrichment of **E**valuations and GAIA stands for **G**raphical **A**nalysis for **I**nteractive **A**id. GAIA was later introduced as a graphical complement to the PROMETHEE rankings.

PROMETHEE method is based on outranking relation between pairs of alternatives on each criterion [13–15]. The PROMETHEE method employs preferential functions to describe the differences between pairs of alternatives. The values of the numerically defined preference functions range from 0 to 1, where zero means no preferential difference between the pair of alternatives and one means absolute preference over the other alternative.

PROMETHEE implements six different predefined preference functions (Usual, U-shape, V-shape, Level, Linear and Gaussian), which may be applied to all quantitative and qualitative attributes with parameters, indifference and preference thresholds [13,16]. Implementation of a decision model with PROMETHEE requires two types of information from the decision maker: weights of the attributes, and type of the preference function.

Scenario1	CLOCK	RAM	Number of c...	Flash	ADC	I2S	SD cards	UART	Support period	YOCTO	LINUX	PRICE	Ethernet	SPI
Unit	MHz	MB	#	MB	#	#	#	#	year	y/n	y/n	Eur	#	#
Cluster/Group														
Preferences														
Min/Max	max	max	max	max	max	max	max	max	max	max	max	min	max	max
Weight	8,79	15,44	2,67	7,56	0,79	5,72	5,92	0,97	12,48	6,46	0,00	20,00	2,67	10,53
Preference Fn.	Linear	V-shape	Usual	V-shape	Usual	Usual	Usual	Usual	Usual	Usual	Usual	Linear	Usual	Usual
Thresholds	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute	absolute
- Q: Indifference	50,00	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	10,00 €	n/a	n/a
- P: Preference	300,00	500,00	n/a	500,00	n/a	n/a	n/a	n/a	n/a	n/a	n/a	50,00 €	n/a	n/a
- S: Gaussian	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Statistics														
Minimum	25,00	64,00	1,00	0,00	0,00	0,00	1,00	1,00	2018,00	0,00	1,00	24,00 €	0,00	1,00
Maximum	1400,00	2048,00	4,00	8192,00	16,00	4,00	4,00	8,00	2028,00	1,00	1,00	490,61 €	2,00	6,00
Average	800,15	637,60	1,85	2250,26	2,79	1,74	2,09	4,23	2026,00	0,70	1,00	124,30 €	1,04	3,21
Standard Dev.	257,65	496,03	1,07	2299,49	3,62	1,29	1,01	1,68	2,70	0,46	0,00	71,20 €	0,33	1,55
Evaluations														
TX6U-8030 (TX6DL)	800,00	1024,00	2,00	4096,00	0,00	2,00	2,00	3,00	2027,00	yes	yes	110,66 €	1,00	1,00
TX6Q-1110 (TX6Q)	1000,00	1024,00	4,00	128,00	0,00	2,00	2,00	3,00	2027,00	yes	yes	164,09 €	1,00	1,00
TXUL-5010 (TX6UL)	528,00	256,00	1,00	4096,00	2,00	2,00	2,00	8,00	2027,00	yes	yes	42,08 €	2,00	4,00
TX6S-8034 (TX6S)	800,00	512,00	1,00	4096,00	0,00	2,00	2,00	3,00	2027,00	yes	yes	76,27 €	1,00	3,00
TX53-8130 (TX53)	800,00	512,00	1,00	128,00	0,00	2,00	2,00	3,00	2027,00	yes	yes	154,99 €	1,00	1,00
TX6Q-1030 (TX6Q)	1000,00	1024,00	4,00	128,00	0,00	2,00	4,00	5,00	2027,00	yes	yes	132,50 €	1,00	5,00
TX6U-8110	800,00	1024,00	2,00	4096,00	0,00	2,00	4,00	3,00	2027,00	yes	yes	124,82 €	1,00	4,00
TX25-4021 (TX25)	400,00	64,00	1,00	128,00	3,00	2,00	2,00	5,00	2027,00	yes	yes	85,03 €	1,00	3,00
TX28-4130 (TX28)	454,00	65,00	1,00	128,00	0,00	0,00	4,00	6,00	2027,00	no	yes	47,23 €	2,00	4,00
TX48-7020	720,00	256,00	1,00	128,00	1,00	1,00	3,00	6,00	2027,00	no	yes	96,32 €	1,00	2,00
Apalis T30 1GB	1400,00	1024,00	4,00	4096,00	0,00	1,00	2,00	5,00	2025,00	yes	yes	146,10 €	1,00	4,00
Apalis iMX6 Quad...	800,00	2048,00	4,00	4096,00	4,00	3,00	3,00	5,00	2028,00	yes	yes	137,20 €	1,00	4,00
Apalis iMX6 Quad...	1000,00	1024,00	4,00	4096,00	4,00	3,00	3,00	5,00	2028,00	yes	yes	112,15 €	1,00	4,00
Apalis iMX6 Dual ...	800,00	1024,00	2,00	4096,00	4,00	3,00	3,00	5,00	2028,00	yes	yes	117,55 €	1,00	4,00
Apalis iMX6 Dual ...	1000,00	512,00	2,00	4096,00	4,00	3,00	3,00	5,00	2028,00	yes	yes	100,05 €	1,00	4,00
Colibri T20 512MB	1000,00	512,00	2,00	1024,00	4,00	4,00	4,00	5,00	2025,00	yes	yes	117,50 €	1,00	6,00
Colibri T20 512M...	1000,00	512,00	2,00	1024,00	4,00	4,00	4,00	5,00	2025,00	yes	yes	148,65 €	1,00	6,00
Colibri T20 256MB	1000,00	256,00	2,00	512,00	4,00	4,00	4,00	5,00	2025,00	yes	yes	106,95 €	1,00	6,00
Colibri T20 256M...	1000,00	256,00	2,00	512,00	4,00	4,00	4,00	5,00	2025,00	yes	yes	129,45 €	1,00	6,00
Colibri iMX6 Dual...	1000,00	512,00	2,00	4096,00	4,00	4,00	3,00	5,00	2028,00	yes	yes	81,10 €	1,00	4,00
Colibri iMX6 Dual...	800,00	512,00	2,00	4096,00	4,00	4,00	3,00	5,00	2028,00	yes	yes	91,20 €	1,00	4,00
Colibri iMX6 Solo ...	1000,00	256,00	1,00	4096,00	4,00	4,00	3,00	5,00	2028,00	yes	yes	71,30 €	1,00	4,00
Colibri iMX6 Solo ...	800,00	256,00	1,00	4096,00	4,00	4,00	3,00	5,00	2028,00	yes	yes	79,95 €	1,00	4,00
Colibri iMX7 Dual ...	1000,00	1024,00	2,00	4096,00	4,00	2,00	2,00	5,00	2027,00	yes	yes	72,80 €	1,00	4,00
Colibri iMX7 Dual ...	1000,00	512,00	2,00	512,00	4,00	2,00	2,00	5,00	2027,00	yes	yes	63,95 €	1,00	4,00
Colibri iMX7 Solo ...	800,00	256,00	1,00	512,00	4,00	2,00	2,00	5,00	2027,00	yes	yes	58,75 €	1,00	4,00
Colibri iMX6ULL 5...	800,00	512,00	1,00	512,00	7,00	3,00	1,00	8,00	2028,00	yes	yes	n/a	1,00	2,00
Colibri iMX6ULL 2...	528,00	256,00	1,00	128,00	8,00	3,00	2,00	8,00	2028,00	yes	yes	n/a	1,00	3,00

Fig. 1. Part of the PROMETHEE evaluation table

The selection of weights and preference functions for the problem in hand is presented in Fig. 1. Geometrical forms in the fourth row (square, circle, diamond) shows which attributes belong to the same group. Next the row denoted with Min/Max tells what are the preferential values of the attribute. Afterwards the weight of the attributes are specified, followed by the used preferential function.

Preference functions for attributes were chosen depending on their data type. Linear preference function was assigned to *CLOCK* and *Price* attributes due to large span of the compared values. V-shape preference function was assigned to *RAM* and *Flash* attributes due to its capability to mark out COMs with the largest RAM/Flash. Using linear and V-shaped preference functions the values of P, the preference threshold and Q, the indifference threshold, have to be assigned.

For example, the values for the attribute *CLOCK* $Q = 50\,\text{MHz}$ and $P = 300\,\text{MHz}$ mean that for clock differences less than 50 MHz the alternatives are regarded as equal, however if the difference is greater than 300 MHz than the alternative with greater value is preferred. The same observation is valid for the attribute *Price*. For the attributes *RAM* and *Flash* the comparison was performed without indifference threshold, hence the choice of V-shape function.

The *Usual* preference function was used for quantitative attribute *YOCTO support* and for qualitative attributes such as *Number of cores*, *ADC*, *I2S*, *SD cards*, *UART*, *SPI*, *Ethernet* and *Support period*. The rationale behind the selection of this function lies in the small number of evaluation levels for these attributes.

Data were evaluated using PROMETHEE II complete ranking or Phi method. The Phi method requires a total preorder (complete ranking without incompatibilities) from the decision maker, based on which the net preference flow is calculated as:

$$\Phi(a) = \Phi^{+}(a) - \Phi^{-}(a) \tag{1}$$

where $\Phi^{+}(a)$ is the positive outranking flow and $\Phi^{-}(a)$ is the negative outranking flow of the alternative a calculated as:

$$\Phi^{+}(a) = \frac{1}{n-1} \sum_{x \in K} \pi(a, x) \tag{2}$$

$$\Phi^{-}(a) = \frac{1}{n-1} \sum_{x \in K} \pi(x, a). \tag{3}$$

In the last two formulas $\pi(x, a)$ expresses the degree of preference of alternative x over a. The degree of preference is calculated as:

$$\pi(a, x) = \sum_{j}^{k} w_j P_j(a, x) \tag{4}$$

where w_j are weights and $P_j(a, x)$ is the associated preference function of a with regard to x.

3.1 Results Obtained with PROMETHEE

The final required user input is the weight selection for particular attribute groups. The selected weights, shown in Fig. 1, express a contradictory claim of having the most processing and communication power for the lowest price. The rationale behind this selection is the following:

- The onboard sensor interfaces, such analog-to-digital converters, often offer only entry level precision and therefore are regarded as unimportant.
- Having good communication interfaces (USART, SPI and I2C) enables interfacing with arbitrary sensors thus increasing flexibility of the platform.

- Support period is important so that the manufacturer guarantees sufficiently long production life-span as well as backward pin compatibility.
- Finally, the requirement for processing power and onboard memory is self-explanatory.

From the row "Weight" in Fig. 1, it is clear that the three most important attributes are *Price*, *RAM* and *Support period*, respectively and the least important attributes are *ADC* (because there is always a possibility to connect an external to the COM, however ADC differ in the number of bits-resolution) and *UART* communication port (transfer speed is slower than SPI). A snapshot of the part of the list is given in Table 2. The result suggest that the first two best COMs are **Apalis iMX6 Quad 1 GB** and **Colibri iMX6 DualLite 512 MB**.

Table 2. Part of the evaluated results

Actions	Phi	Phi+	Phi−
Apalis iMX6 Quad 1 GB	0.4007	0.5074	0.1067
Colibri iMX6 DualLite 512 MB	0.3652	0.4869	0.1217
Apalis iMX6 Quad 2 GB (IT)	0.3595	0.5044	0.1449
Colibri iMX7 Dual 1 GB	0.3574	0.4793	0.1219
Apalis iMX6 Dual 1 GB (IT)	0.3340	0.4704	0.1365
Colibri iMX6 Solo 256 MB	0.3104	0.4537	0.1433
Apalis iMX6 Dual 512 MB	0.3065	0.4557	0.1492

After excluding a couple of alternatives due to their high prices, low performances and had almost no chances of becoming the best possible alternative, a GAIA plot was created [17]. It is presented in Fig. 2 which provides descriptive complement to the PROMETHEE rankings. Before the evaluation of the GAIA plot the attributes were grouped as described at the end of the Sect. 2. After evaluation of the GAIA plot, the best possible alternatives lays in direction of the decision vector denoted with the red thick line, which is calculated using the Principal Component Analysis method over all attribute values. In addition to that the quality of the evaluated GAIA plot in (U, V) plane is 73% [17]. On the figure it can also be seen that *Price* vector is denoted with red thin line with a square at the top, *CPU* vector is denoted with pink thin line with a circle at the top, *ADC* vector is denoted with dark blue thin line with a diamond at the top, *Support* vector is denoted with light green thin line with a circle at the top and *Communication* vector is denoted with light blue thin line with a diamond at the top. An interesting thing to mention is that the Raspberry PI is on the opposite site of the decision vector and has very little chances to be become a best possible alternative (according our preferences) even though is the cheapest and most well-known alternative.

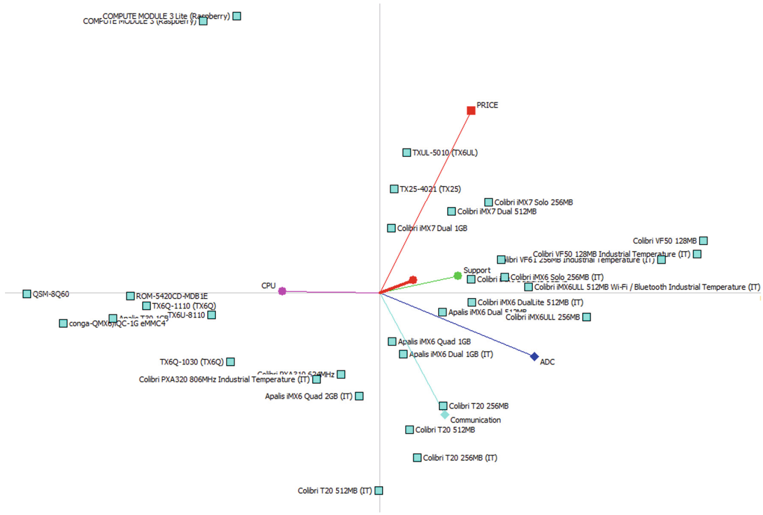


Fig. 2. GAIA presentation of the PROMETHEE rankings (Color figure online)

4 Modelling with DEX Method

DEX method defers from PROMETHEE in several ways. Firstly, DEX belongs to the group of qualitative MCDM methods. Secondly, decision rules in DEX are provided by the decision maker in a form of “if-then” rules. In DEX, the qualitative attributes build a hierarchical structure which represents a decomposition of the decision problem into smaller, less complex and possibly easier to solve sub-problems [18,19].

There are two types of attributes in DEX: basic attributes and aggregated ones. The former are the directly measurable attributes, also called input attributes, that are used for describing the options. The latter are obtained by aggregating the basic and/or other aggregated attributes. They represent the evaluations of the options. The hierarchical structure in DEX represents a tree. In the tree, attributes are structured so that there is only one path from each aggregate attribute to the root of the tree. The higher-level attributes depend on their immediate descendants in the tree. This dependency is defined by a utility function. The higher-level attribute, its immediate descendants and the utility function form a qualitative decision table. The decision maker’s preferences over the available options are given with the attribute that is called a qualitative class, or only a class. Options that are almost equally preferred belong to the same qualitative class.

Within DEX, which is implemented as DEXi [20], the decision maker is requested to define a qualitative multi-attribute model. The DEXi model consists of:

Attributes: variables that represent basic features.

Scales of attributes: these are qualitative and consist of a set of words, such as: 'excellent', 'acceptable', etc. Usually, scales are ordered preferentially, i.e., from less preferred to most preferred values.

Hierarchy of attributes: decomposition of the decision problem and relations between attributes; higher-level attributes depend on lower-level ones.

Decision rules: tabular representation of a mapping from lower-level attributes to higher-level ones. The table is called utility function and specifies all values of the higher-level attribute as a combination of all values of its lower-level attributes.

The same problem from above was used for modelling a decision support system in DEXi. The model structure shown in Table 3 consists of 13 different attributes. DEX is a methodology that deals with qualitative attributes, hence a predefined qualitative values have to be assigned for the quantitative ones. For each of the aggregated attributes in Table 3 (Serial, Communication, Memory, Speed, CPU, System, Support, S+R[2] and System on Module), a tabular utility function is defined by the decision maker. The weights are set automatically from the utility functions, however the program allows the user to set the weights manually as well.

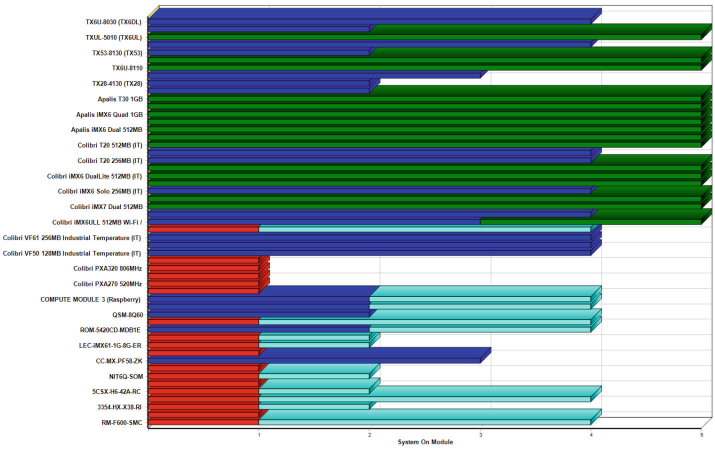


Fig. 3. Results from the evaluation in DEXi (Color figure online)

4.1 Results of Modeling with DEX

DEX evaluation results are discrete and belong into several ordered categories instead of obtaining a fully ranked alternatives. Alternatives which belong to the same category are considered as indistinguishable (Fig. 3). Alternatives that belong to the most preferred class are given in Table 4.

[2] S+R stands for **S**ystem and Suppo**r**t.

Table 3. DEXi model tree and attribute scales for assessment of COMs

Attribute	Scale
SISTEM ON MODULE	1, 2,3,4, *5*
— Price	very high, high, medium, low, *very low*
— S+R	1, 2,3,4, *5*
—— SUPPORT	1, 2,3,4, *5*
——— ProductionPeriod	1, 2,3,4, *5*
——— Yokto	short, medium, *long*
—— SISTEM	no, *yes*
——— CPU	1, 2,3,4, *5*
———— SPEED	1, 2,3,4, *5*
————— Number of cores	1, 2,3,4, *5*
————— Clock	1, 2, *4*
———— MEMORY	1, 2,3,4,5,6, *7*
————— Flash	1, 2,3,4, *5*
————— RAM	1, 2,3,4, *5*
——— COMMUNICATION	1, 2,3,4, *5*
———— SDCards	1, 2,3,4, *5*
———— Ethrenet	1, 2,3, *4*
———— SERIAL	0, 1, *2*
————— I2S	1, 2,3,4, *5*
————— SPI	<2, 2, *>2*
————— UART	<2, 2, *>2*
——— ADC	0, 1, *>1*

Table 4. COMs with highest mark according to DEXi evaluation

COM	DEXi mark
TXUL-5010 (TX6UL)	5
TX6Q-1030 (TX6Q)	5
TX6U-8110	5
Apalis T30 1 GB	5
Apalis iMX6 Quad 2 GB (IT)	5
Apalis iMX6 Dual 1 GB (IT)	5
Apalis iMX6 Dual 512 MB	5

All rankings of COMs are visually presented in Fig. 3. Bars that consists of two or more colors means that some attribute parameter is missing. For example, for the alternative **Colibri iMX6ULL 512 MB Wi-Fi/Bluetooth** the *Price* is not defined and hence the evaluation for this alternative is between 3 and 5, which depends of the *Price* value.

5 Discussion and Conclusion

The selected alternatives are a representative sample of the currently available COMs spanning a broad price range. There are two conclusions from the obtained results. First, the most popular COM with the general public, the Raspberry PI, belongs in the middle of the preference scale. Despite the low price, this COM lacks clear information regarding the product longevity support. On the other hand, the top choices, the Apalis and Colibri families of COMs, are tailored made for industrial applications. As a result, both families are rated for industrial environments (proper certification), have predetermined pin compatibility life span, allow connectivity of plethora of sensors and most importantly support the latest operating systems. These positive points outweigh the higher price, thus making them the optimal choice.

In this paper two models for selection of Computer on Modules ware developed using two different methods: PROMETHEE and DEX. These methods were chosen for building a DSS for selection of the most appropriate COMs, because they are both implemented into freely accessible tools, are easy to use and provide a sufficient documentation. Both methods are well known and are widely used for providing decision support in various areas.

PROMETHEE belongs to the group of outranking methods, and it is designed to handle both qualitative and continuous quantitative data, while DEX is mainly used for modeling qualitative data, using rule based "if-then" system. When compared, both programs provide consistent results regarding the selection of the most appropriate COMs in terms of assigning the best alternatives to the highest preferential category. PROMETHEE implements the total ranking of the alternatives, thus leading to an easy choice of the best alternatives, while DEX provides a group of best COMs and an understandable explanation of how they differ. For example the best solution according to the model obtained with PROMETHEE is Apalis iMX6 Quad 1GB, where DEXi for the same COM evaluates with the highest class of 5 together with 12 other COMs evaluated to belong to the same class.

References

1. Gerber, A.: Choosing the best hardware for your next IoT project, May 2017. https://www.ibm.com/developerworks/library/iot-lp101-best-hardware-devices-iot-project/iot-lp101-best-hardware-devices-iot-project-pdf.pdf. Accessed 14 Feb 2018
2. Maksimović, M., Vujović, V., Davidović, N., Milošević, V., Perišić, B.: Raspberry Pi as internet of things hardware: performances and constraints. Des. Issues **3**, 8 (2014)
3. Polianytsia, A., Starkova, O., Herasymenko, K.: Survey of hardware IoT platforms. In: 2016 Third International Scientific-Practical Conference on Problems of Infocommunications Science and Technology (PIC ST), pp. 152–153. IEEE (2016)
4. Pitchipoo, P., Venkumar, P., Rajakarunakaran, S.: Modeling and development of a decision support system for supplier selection in the process industry. J. Ind. Eng. Int. **9**, 23 (2013)

5. Chan, F.T.S., Chan, H.K., Ip, R.W.L., Lau, H.C.W.: A decision support system for supplier selection in the airline industry. Proc. Inst. Mech. Eng. Part B: J. Eng. Manuf. **221**, 741–758 (2007)
6. Dweiri, F., Kumar, S., Khan, S.A., Jain, V.: Designing an integrated AHP based decision support system for supplier selection in automotive industry. Expert Syst. Appl. **62**(Suppl C), 273–283 (2016)
7. Kar, A.K.: Revisiting the supplier selection problem: an integrated approach for group decision support. Expert Syst. Appl. **41**(6), 2762–2771 (2014)
8. Chen, D., Cong, J., Gurumani, S., Hwu, W., Rupnow, K., Zhang, Z.: Platform choices and design demands for IoT platforms: cost, power, and performance tradeoffs. IET Cyber-Phys. Syst.: Theory Appl. **1**(1), 70–77 (2016)
9. Samie, F., Bauer, L., Henkel, J.: IoT technologies for embedded computing: a survey. In: Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis, p. 8. ACM (2016)
10. Promethee-Gaia: Get Started with Visual PROMETHEE. http://www.promethee-gaia.net/documents.html
11. Bohanec, M.: DEXi: Program for Multi-Attribute Decision Making, User's Manual, Version 5.00. IJS report DP-11897. Jožef Stefan Institute, Ljubljana (2015). http://kt.ijs.si/MarkoBohanec/DEXi/html/DEXiDoc.htm
12. Philips Semiconductors: I2S bus specification, June 1996
13. Brans, J.-P., Vincke, P., Mareschal, B.: How to select and how to rank projects: the Promethee method. Eur. J. Oper. Res. **24**(2), 228–238 (1986)
14. Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdasi, M.: Promethee: a comprehensive literature review on methodologies and applications. Eur. J. Oper. Res. **200**(1), 198–215 (2010)
15. Brans, J.-P., Mareschal, B.: Promethee methods. In: Figueira, J., Greco, S., Ehrgott, M. (eds.) Multiple Criteria Decision Analysis, State of the Art Survey. Springer Science, New York (2005). https://doi.org/10.1007/0-387-23081-5_5
16. Zhaoxu, S., Min, H.: Multi-criteria decision making based on Promethee method. In: 2010 International Conference on Computing, Control and Industrial Engineering (CCIE), vol. 1, pp. 416–418. IEEE (2010)
17. Mareschal, B., Brans, J.-P.: Geometrical representations for MCDA. Eur. J. Oper. Res. **34**(1), 69–77 (1988)
18. Bohanec, M., Žnidaršič, M., Rajkovič, V., Bratko, I., Zupan, B.: DEX methodology: three decades of qualitative multi-attribute modeling. Informatica **37**(1), 49–54 (2013)
19. Bohanec, M.: Decision making: a computer-science and information-technology viewpoint. Interdiscip. Descr. Complex Syst. **7**(2), 22–37 (2009)
20. Bohanec, M.: DEXi: Program for Multi Attribute Decision Making, User's Manual. Jožef Stefan Institute, July 2015

Unveiling Hidden Patterns in Flexible Medical Treatment Processes – A Process Mining Case Study

Kathrin Kirchner[1(✉)] and Petar Marković[2]

[1] Berlin School of Economics and Law, 10315 Berlin, Germany
kathrin.kirchner@hwr-berlin.de
[2] Technische Universiteit Eindhoven, 5600 Eindhoven, The Netherlands
p.markovic@tue.nl

Abstract. In hospital environments, treatment processes, resp. clinical pathways, are adopted based on the health state of a patient. Modeling of pathways is time consuming and due to the involvement of many participants, the introduction of clinical pathways is cost-intensive. Process mining offers a possibility for automatic or semi-automatic creation of clinical pathways based on the event log data recorded during the process execution in hospital information systems. However, state-of-the-art algorithms struggle to discover meaningful end-to-end patterns from highly flexible clinical log data. This challenge can be addressed by Local Process Models. They allow pathways to be modeled partially, thus enabling the detection of major process steps. In our case study, we apply this recently proposed method on a real world clinical dataset and discuss results and challenges.

Keywords: Process mining · Local process model · Clinical pathway
Flexible process · Case study

1 Introduction

Healthcare faces the challenge to deliver high treatment quality and patient satisfaction while being cost efficient. Clinical pathways as structured, multidisciplinary care plans standardize treatment processes. They define the steps of patient care for a certain disease in a specific hospital [16]. Compared to processes in industry, clinical pathways are more flexible as a treatment process varies for each individual patient. Additional therapies might be necessary and the sequence of treatment steps might change due to interpreting patient-specific data. Thus, deviations from pre-planned treatment processes are frequent, and their effect on the process is much stronger. This is especially evident in certain medical domains, such as organ transplantation, where treatment processes can last long and the unstructuredness is high.

Data about the treatment of a patient is collected in different clinical information systems in a hospital. Treatment steps (events) can be recorded based

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 169–180, 2018.
https://doi.org/10.1007/978-3-319-90315-6_14

on timestamps and each event refers to a particular patient and a particular activity. More information might be available, for example, the performer of the event, the timestamp or documents or treatment information recorded with the event. This data can be collected in an event log. The idea of process mining is to discover, monitor and improve real processes based on the data in event logs. [22] In this paper the focus is on process discovery where no a-priori process schema exists. Based on an event log, a process mining algorithm constructs a schema of the process.

Traditionally, the majority of the state-of-the-art process mining algorithms search for complete, so called "start-to-end" processes. However, in highly flexible settings, such as the one this paper addresses, they fail to achieve proper results, or appear to have limited effectiveness. This fact is supported by the previous study of [7], where the robust trace clustering approach of [3] was only partial successfully applied on living donor liver transplantation data, the same dataset this paper is referring to. Although one cluster with patients following a typical treatment process could be found, the methodology was not able to find useful pathways for the remaining untypical data. The case study faced challenges because of coarse timestamps, small size of sample, and different lengths of the pathway for different patients.

A recent approach, Local Process Models (LPM) Discovery, is trying to answer these challenges. It is focused on the mining of a set of process models where each model describes the behavior represented in the event log only partially, i.e., local process models are created [20]. This can be especially helpful if no overall process model exists that describes the traces of the process from the start of a process to the end.

In this paper, we use LPMs for discovering short process sequences in a case study of living liver donor transplantation. These sequences represent interpretable derived sub-processes, potentially useful for building models on a higher level of abstraction. The paper is structured as follows: The next chapter discusses shortly the related work. Section 3 explains the used data and methodology. Section 4 shows our results, and a summary and outlook is given in Sect. 5.

2 Selected Related Work

2.1 Process Mining in Healthcare

Process mining follows the aim to extract novel patterns from event data. Process discovery as a part of process mining focuses on discovering a process model for the set of event log sequences from start to end of a process [22]. In many cases, events recorded in the event log are too fine-grained, causing process discovery algorithms to discover incomprehensible process models or process models that are not representative for the event log [19]. This might lead to an uninterpretable mess of nodes and connections in the process model - a so called spaghetti model.

A literature review of the application of process mining in healthcare is given by [15]. It was found that the most commonly used techniques are Trace Clustering [17], Fuzzy Miner [4] and Heuristics Miner [25], because they can manage noise and incompleteness, and allow models to be identified for less-structured processes. Furthermore, they allow similar cases to be grouped together, as is the case of trace clustering. Process Mining was applied in 22 different medical fields with oncology and surgery being the most prominent ones. As an example, [14] applied heuristics miner on log data from dentistry and described difficulties to handle flexibilities in the clinical pathways. [6] List challenges for process mining in healthcare, among them complexity and flexibility of treatment processes and that not all treatment steps are fully recorded in a clinical information system.

2.2 Local Process Models

To overcome the problem of obtaining spaghetti processes, a group of techniques exists that is able to reveal stronger patterns of local character (i.e. smaller pieces of traces consisting only of neighboring activities) in the process. Contrary to global approaches (e.g., [10]), these algorithms provide useful insights into the process structure, since these detected partial patterns could be observed as structured steps/parts of the flexible processes on a higher level of abstraction. These techniques are quite alike to frequent sequence mining [1] and use a similar logic.

One of the techniques identified as especially appropriate for addressing the problem is the novel Local Process Models methodology proposed by [20]. Its nature is similar to the very mature Episode mining [9], yet, with many additional advantages presented in Table 1. In this table we compare LPM and Episode Mining with general global process mining techniques.

For our case study, we selected the LPM approach, because it provides the ability to discover processes that are formal and include all possible kinds of behavior (sequences, concurrency, choices, loops). Furthermore, the existence of a transparent, highly customizable set of quality measures was the key reason for evaluating LPM on our data. Due to space restrictions, only a brief formal introduction of LPMs is given, whilst for the in-depth explanation of the method, readers should refer to [18] or [20].

Formally, Local Process Models (LPMs in the further text) are Process Trees [2] of size (i.e. number of leaf nodes) between 2 and 5, built of frequently re-occurring activities originating from event log subtraces. Process trees consist of leaf nodes representing activities, and non-leaf nodes representing relation operators between activities:

- $\rightarrow$ sequential execution: second child executed after first,
- $\wedge$ parallel execution: first and second child are both executed in any order,
- $\times$ exclusive choice: either first or second child is executed,
- $\circlearrowleft$ loop: after the execution of the first, *"do"* child, the second, *"re-do"* child, followed again by "do" child, could be executed afterward, minimally once.

The composition of two process trees is also a process tree. For example, given a *language* (i.e. the set of allowed activity execution paths) of the example model, $\mathcal{L}(M) = \{\langle b, c, a\rangle, \langle b, a, c\rangle, \langle d, a, c\rangle, \langle d, c, a\rangle\}$, a resulting process tree M has a

Table 1. Characteristics of LPM, Episode Mining and global techniques

Local process models [20]	Episode miner [9]	Global techniques, e.g. [10]
Advantages		
Able to mine frequent process steps in flexible logs (local scope)	Able to mine frequent process steps in flexible logs (local scope)	Large number of mature, scalable and robust techniques to choose from
Able to capture all of process behavior (sequential, parallel, loops and exclusive choice flow)	Computationally efficient for larger event logs	
Formal process models based on process trees, thus sound by design	Informal process models, thus unstable results	
Represented as Petri nets, thus easily convertible to other notations		
Large number of adjustable quality metrics		
Disadvantages		
Computationally inefficient (long running times for large sized logs w.r.t. number of distinct activities), but has heuristics for speed-up	Limited type of process behavior can be modeled (only sequential and parallel flow, no loops or exclusive choice)	Observes process as a whole, from "Start-to-end" without possibility to focus on relatively stable process parts (Global scope)
		Limited quality metrics
		Proven extremely low performance on flexible processes
		Many of the techniques are computationally over-complex

form of $\rightarrow(\times(b,d), \wedge(a,c))$, as first b or d is executed, followed by a and c, in any order.

In LPMs, underlying process tree models are represented as Accepting Petri-nets $APN = (LPN, M_o, MF)$, that is, labeled Petri-nets with an initial marking $M_o \in \mathbb{N}^p$ and a set of possible final markings $MF \subseteq \mathbb{N}^p$, such that $\forall_{M_1,M_2 \in MF}\ M_1 \nsubseteq M_2$, where a marking represents a state of a Petri net (i.e. distribution of tokens over its places). A labeled Petri net is defined as a bipartite graph, i.e. a tuple $LPN = \langle P, T, F, \Sigma_M, \ell \rangle$ consisting of a set of places P, a set of transitions (activities) T such that $P \cap T = \emptyset$, and arcs (flow-relation $F = (P \times T)(T \times P)$). The set of labels representing the names of activities is Σ_M (with invisible events $\tau \notin \Sigma_M$), and the labeling function $\ell : T \rightarrow \Sigma_M \cup \{\tau\}$).

The intention of the approach is to partition the event log L, consisting of the traces $\sigma = \langle e_1, e_2, \ldots, e_n \rangle \in L$ (sequences of events) in the way that the number of events that would fit the language-fitting LPM subtrace sequences is maximized; i.e. a segmentation of a LPM trace σ: $\lambda(LPM) = \alpha_0\beta_1\alpha_1\beta_1 \ldots \beta_n\alpha_n$ where $\beta_j \in \mathcal{L}(LPM)$ represents a language-fitting subsequence, whilst $\alpha_j \notin \mathcal{L}(LPM)$ is a non-fitting subsequence, such that the number of events in $\{\beta_1, \beta_2, \ldots, \beta_n\}$ is maximized [20]. E.g., for a given trace $\sigma = \langle b, a, c, d, a, d, b, d, c, a, b, c, b \rangle$ projected on activity set $\{b, c, d\}$, an optimal partition, so that fitting subtraces are $\beta_1 = \langle b, c, d \rangle$ and $\beta_2 = \langle d, b, c \rangle$ (with corresponding non-fitting subtraces $\alpha_0 = \langle\rangle, \alpha_1 = \langle d \rangle$ and $\alpha_3 = \langle b, c, b \rangle$) results in a projected trace $\sigma' = \langle b, c, d, d, b, c \rangle$.

LPM discovery has four repetitive steps. In the first step, a pool of models fitting to the partitioning is *generated*, which are then *evaluated*, and the best ones *selected* and kept, whilst the others are disregarded. In the final fourth step, the best candidate models are *expanded*, i.e. their activity nodes are replaced by operator nodes whose children are the replaced activity node and another node from the language. The cycle is repeated until there is no tree in the candidate set that meets the desired threshold. During the detection, in the evaluation step, LPMs are being evaluated using the weighted average of the following quality measures [21] (formulas omitted due to space restrictions):

1. *Support* - the frequency of the detected LPM in the log (number of detected fragments in the log). The higher the support, the stronger the model.
2. *Confidence* - the percent of the events of the activities in the log that abide to LPM. The higher the confidence, the stronger the model.
3. *Language fit* - ratio of the behavior allowed by the LPM that is observed in the event log. LPMs that allow more behavior than it is observed have tendency to overgeneralize.
4. *Determinism* - the average number of enabled transitions in each marking of the model that was reached while replaying the event log. LPM that has higher non-determinism is giving less information on ordering of the events (i.e., models with high level of non-determinism are so called "flower-models"). The higher the value, the better the model.
5. *Coverage* - The frequency of activities described in the LPM in the event log. The higher the value, the better the model.

These metrics represent local equivalents of global quality metrics defined in [23]. For different types of problems, different measures (and their corresponding weights) can be preferred in order to detect models of the best quality in terms of "local" fitness, precision and generalization (while simplicity is controlled with input size of the model), in a particular case. Since this is a domain-related aspect, it will be investigated in more detail in the results section later. Usually, a trade-off between quality dimensions is needed, since a perfect model does not exist. However, in our study, besides objectively good models, we also prefer the ones with the highest level of interpretability.

In order to overcome high computational complexity of the full-sized generation of subsets of activities in the a-priori-like manner, and enable the analysis of the larger logs, [21] introduced a set of three heuristics (Markov clustering, Log entropy, and Maximal relative information gain) which, in addition to the execution speedup, simultaneously allow mining of higher quality models through log projections. The authors have reported that the selection of an appropriate heuristic is highly dependent on the specific log and domain.

Finally, the most promising characteristic in the LPM methodology is its ability to use the detected patterns for generating the abstractions of process steps as shown in [19]. Thus, we applied this approach on our data set. We argue that this way, we are able to detect relatively stable subprocesses in the flexible process. These low level patterns can be abstracted to a high-level log [12] which can be mined using global process mining techniques [13].

3 Data and Method

During the research project PIGE (Process Intelligence in Healthcare), a clinical pathway for living liver donors was modeled in BPMN together with medical personnel [5]. The process can be roughly described as follows (BPMN diagram in Fig. 1): A healthy person can donate a part of her/his liver to a near relative. First, medical doctors have a first talk and investigation with all possible donors for a certain patient. The preselected person, before he/she can become a living donor, has to undergo an evaluation procedure to ensure that the individual is physically fit. Computer tomography (CT) scans or magnetic resonance tomography (MRT) are done to image the liver. The pre-examinations are predetermined, but can change in the sequence depending on the availability of necessary resources. During and after operation, complications can occur that lead to additional interventions or even an additional operation.

Fig. 1. BPMN process model of living liver donor (high level)

Besides of the high-level process shown in Fig. 1, 10 subprocesses were modeled on three more detailed levels. Flexibility in process execution was expressed using, e.g., annotations with a list of possible steps or documents named *checklist.* [8] Especially for the *evaluation of potential donor* (Fig. 1, a checklist is used that lists all mandatory and possible additional investigations for a potential donor. The sequence of these investigations are planned by the medical personnel due to availability of resources like medical experts and tools. The interesting question occurs whether hidden patterns in the treatment execution data can be unveiled that can give interesting insights in the treatment process: Which treatment steps are more often conducted in a direct sequence than others? This might be helpful in a hospital to optimize the treatment procedures, which is in our case especially the pre-evaluation subprocess before the operation.

The data set was extracted from a clinical information system during the PIGE project. All patient data which were marked as living liver donors in a time period of 3 years were selected and anonymized. The resulting data set contained 50 living liver donors with all together 331 events. Not all patients went through all process steps. If the pre-examination found the person not suitable for donating the liver, an operation is not done. Therefore, the number of process steps for patients was different. Patients that were already in a later process step in the considered time period were also in the data set. Thus, not all pathways had the same start- and endpoint. Furthermore, the timestamps for all events were only dates (no concrete time was available), and thus, several events were executed on the same day (Table 2). The event log file consists of treatment steps on a higher level of detail (e.g., CT as a part of the evaluation procedure).

Table 2. Event log file for living liver donors

Patient-ID	OPS-code	Treatment	Treatment day	Admission day	Discharge day
12345678	3-225	CT:Abdomen	10.10.2014	10.10.2014	12.10.2014
12345678	3-226	CT:Pelvis	10.10.2014	10.10.2014	12.10.2014
...					
23456789	3-225	CT:Abdomen	08.02.2015	08.02.2015	10.02.2015

The average trace has length of approximately 7 events. There are, on average, 2.07 ± 1.52 activities per day. 32 distinct activities can be found in the log. Figure 2 depicts the relative distributions of recorded activities per day. The figure straightforwardly show a negligible negative effect of timestamp granularity to the future conclusions, since the majority (cca 72%) of the population has no more than two activities executed on the same day.

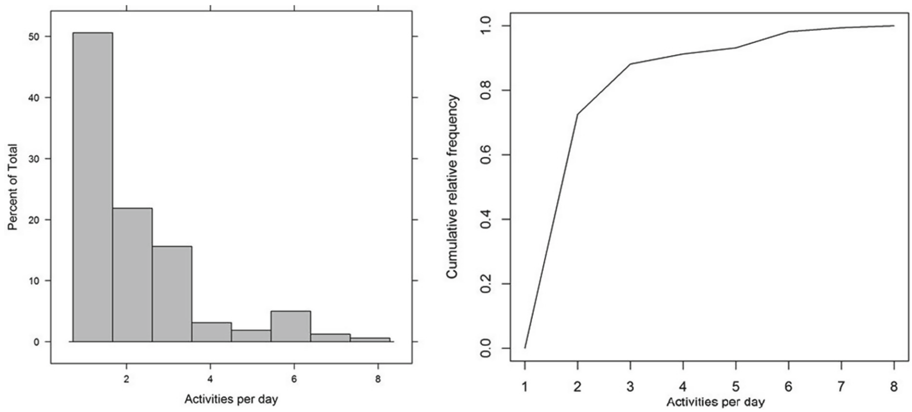


Fig. 2. Relative and cumulative distribution of activities per day - donors

Taking all previously mentioned into account, the methodology we used in this paper, consisted of the following four steps:

1. Mine for k-sized Local Process Model (where k can take values from 2 to 5) with non-restrictive settings of a-priori thresholds
2. Examine the obtained models and select the best ones, according to the domain knowledge and quality measures weights most appropriate for the purpose.

3. Repeat steps 1–2 with application of heuristics
4. When the best LPMs are known, use them for inducing the higher level of abstraction on the non-flexible process steps, so that the process can be analyzed from the higher level.

4 Results

In order to find local process models, we used ProM process mining software [24] with the LPM algorithm plugin. We intended to mine for LPMs of sizes fitting the full possible range (i.e. models consisting of 2, 3, 4, and 5 activities). However, the algorithm was able to detect only one model consisting of more than 4 tasks. Altogether, we calculated 60 different models using the following weights for support: 0.1, language fit: 0.1, confidence: 0.4, coverage: 0, determinism: 0.3 and average number of firings: 0.1. For finding LPMs, the traditional approach (no heuristics used) was applied. We found 18 models with 2 tasks, 27 with 3 tasks and 14 comprising 4 tasks. Figure 3 gives examples of five mined LPMs from the living liver donors.

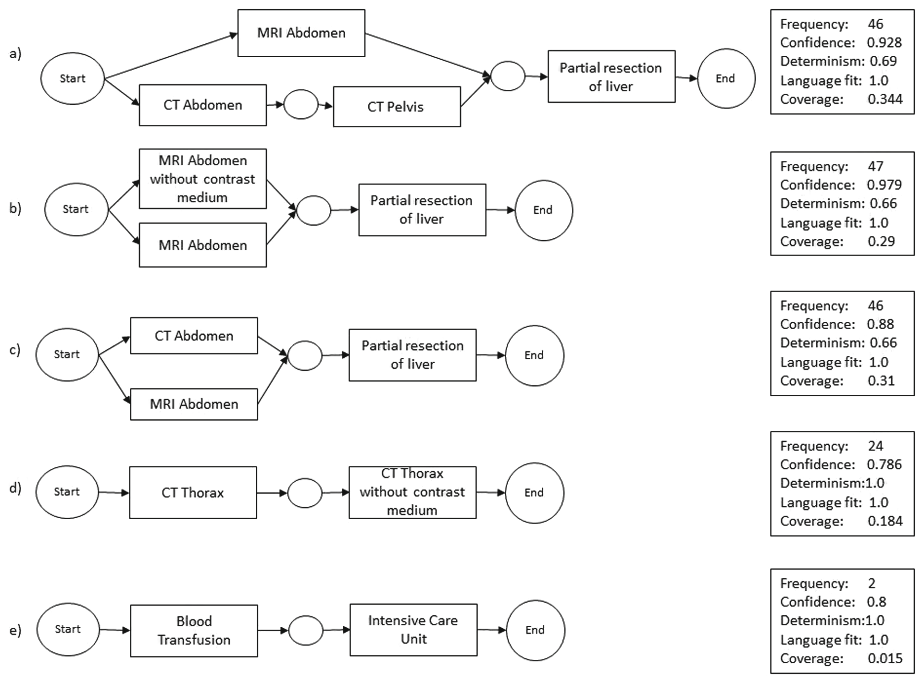


Fig. 3. Examples of mined local process models

Figure 3(a) shows the only one relevant model with four process tasks, that was valid for 46 cases. All other 13 mined process models with four tasks were only valid for 2 patients. In the visualized model, the patient undergoes an evaluation before the liver operation. In the evaluation, most patients get a MRI, while some get alternatively a computer tomography of abdomen and pelvis. Figure 3(b) and (c) show also evaluation steps before the operation can be done. In (d) two CTs are executed. The first CT does not provide useful results, so a second CT using contrast medium is done. In (e), that is only valid for two patients, the post-operation phase is visible: First, the patient gets a blood transfusion, before he is brought to intensive care unit.

The mined local process models define high-level tasks such as: ≪Evaluation≫, ≪Operation≫ or ≪Post-operative Phase≫, which are useful for raising the level of abstraction of the process, and hence, its interpretability (see also Fig. 1):

- Figure 3a–c: Evaluation of a potential donor → Operation of a living liver donor
- Figure 3d: Evaluation of a potential donor
- Figure 3e: Operation of a living liver donor → Post-Operative treatment in the hospital

The 60 mined local models have different frequencies. In the case of models containing 2 steps (like in Fig. 3e), we found frequent patterns with frequency 44 or 37 (total number of patients: 50), but also two infrequent patterns that only occurred for 2 patients. In case of patterns that were 3 steps long, we found 10 patterns that are valid for 44 out of 50 patients.

Our case study faced some challenges, because of the coarse timestamps (date instead of date and time), the small size of sample (only 50 patients), and the large deviation of the pathway length. Nevertheless, we could identify interesting local process models. Even if the frequency of a model is low, it might provide interesting insights for medical personnel (e.g., handling complications during operations). An additional issue was that not all process steps were recorded in the clinical information system, e.g., the step *admission to hospital for operation*.

In addition to the data set used in this case study, we investigated another event log of liver transplantation patients, consisting of 2294 events referring to 256 cases in the period of 586 recorded days. The average trace consists of approximately 9 events. On average, there are 3.91 ± 3.08 activities per day (Fig. 4). In terms of the distribution of activities per day, this log is more complex and sensitive to the timestamp granularity issue than the living liver donor data, since less than 50% of the patients has up to two activities executed on the same day. Nevertheless, due to the highly skewed distribution (indicated with high standard deviation) and the nature of the LPM method, it could be presumed that the negative effects would be minimal. Unfortunately, in this case, there are 234 distinct activities, which is currently too large and complex to be analyzed due to computational limitations of the LPM methodology. Thus, we applied all the heuristics proposed in [21] in addition to the traditional LPM approach. However, the results showed that the log size limitations for application of these heuristics (even the most advanced Markov clustering one), still were unable to overcome this issue.

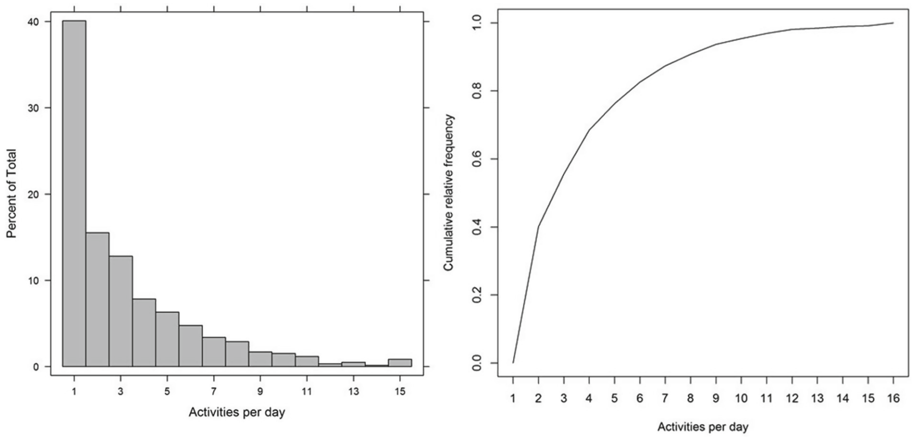


Fig. 4. Relative and cumulative distribution of activities per day - liver transplantation patients

Regardless of the faced issues, it should be stressed that the results bring valuable insights, as to the best of the authors' knowledge, there exists only one paper where LPMs were applied on medical data [13] until now. In this paper, a large pool of LPMs used for abstracting event logs is generated in automatic fashion. However, interpretability and usefulness of generated models is not discussed, nor the models and their corresponding settings are presented, as the main aim of the paper is overall benchmark assessment of the robustness of the approach, which is solely based on aggregated quality metrics discussion (i.e. F-score values change compared to the non-abstracted model), and not the analysis of the underlying medical process. In addition, LPMs were mined on a significantly less flexible and larger event log describing patients with SEPSIS [11], originating from hospital ERP system, consisting of approx. 1000 cases, 15000 events, and only 16 distinct activities. The log was already successfully mined using end-to-end approaches beforehand, thus although the algorithm possibly found interesting process steps, it is debatable whether the discovered knowledge yielded any new previously unknown valuable insights about the process for decision makers.

5 Summary and Outlook

In this paper, we provided a case study on how flexible event logs from medical domain can be successfully mined and analyzed within a local scope using LPMs. Contrary to the complicated end-to-end spaghetti models, the local process models allow for better understanding of major process steps, thus improving communication of results with medical personnel. Due to their simple structure and high interpretability, they allow the validation of the results using the expert knowledge, which is of utter importance for the specific domain. The naming of the local process models has to be done by a medical expert. Discovered local process models are used for raising the level of abstraction into a single subprocess step. Thus, the flexible parts of the treatment process are explained using local process models.

Although the LPM methodology provided good results on the living liver donor data, it was not possible to use it on the larger data set of liver transplantation patients. Thus, in the future, a smart preprocessing of the liver transplantation patients' dataset is intended in order to enable execution of LPM miner on it. Since there exist a large number of activities which are recorded on the level of granularity lower than needed (e.g. different types of CTs are recorded separately), the log size could be reduced without large loss of information (i.e. it is important to know that CT activity is done, whilst it is not important to know which specific type was applied). Successful implementation of this improvement could lead to more convincing and stable findings.

Acknowledgment. The authors kindly thank Niek Tax, Ph.D. candidate at TU Eindhoven, the author of the LPM methodology, for his valuable help and advice on installing and running the novel LPM mining ProM nightly build plug-in, which was crucial for conducting the experimental part of this paper.

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of the 11th International Conference on Data Engineering, pp. 3–14. IEEE (1995)
2. Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P.: A genetic algorithm for discovering process trees. In: IEEE Congress on Evolutionary Computation (CEC), pp. 1–8. IEEE (2012)
3. Delias, P., Doumpos, M., Grigoroudis, E., Manolitzas, P., Matsatsinis, N.: Supporting healthcare management decisions via robust clustering of event logs. Knowl.-Based Syst. **84**, 203–213 (2015)
4. Günther, C.W., van der Aalst, W.M.P.: Fuzzy mining – adaptive process simplification based on multi-perspective metrics. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) BPM 2007. LNCS, vol. 4714, pp. 328–343. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75183-0_24
5. Herzberg, N., Kirchner, K., Weske, M.: Modeling and monitoring variability in hospital treatments: a scenario using CMMN. In: Fournier, F., Mendling, J. (eds.) BPM 2014. LNBIP, vol. 202, pp. 3–15. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-15895-2_1
6. Homayounfar, P.: Process mining challenges in hospital information systems. In: Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1135–1140. IEEE (2012)
7. Kirchner, K., Marković, P., Delias, P.: Challenges in automatic creation of clinical pathways: a case study. In: Proceeding of XV International Symposium SYMORG 2016, pp. 188–194 (2016)
8. Laue, R., Kirchner, K.: Using patterns for communicating about flexible processes. In: Conference on Business Process Modeling, Development, and Support (BPMDS), pp. 12–19. CEUR (2017)
9. Leemans, M., van der Aalst, W.M.P.: Discovery of frequent episodes in event logs. In: Ceravolo, P., Russo, B., Accorsi, R. (eds.) SIMPDA 2014. LNBIP, vol. 237, pp. 1–31. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-27243-6_1

10. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Discovering block-structured process models from event logs - a constructive approach. In: Colom, J.-M., Desel, J. (eds.) PETRI NETS 2013. LNCS, vol. 7927, pp. 311–329. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38697-8_17
11. Mannhardt, F.: Sepsis cases - event log. Eindhoven University of Technology (2016). https://doi.org/10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460
12. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P., Toussaint, P.J.: From low-level events to activities - a pattern-based approach. In: La Rosa, M., Loos, P., Pastor, O. (eds.) BPM 2016. LNCS, vol. 9850, pp. 125–141. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45348-4_8
13. Mannhardt, F., Tax, N.: Unsupervised event abstraction using pattern abstraction and local process models. In: Proceedings of the Business Process Modeling, Development, and Support Working Conference 2017 (BPMDS) (2017)
14. Mans, R., Reijers, H., van Genuchten, M., Wismeijer, D.: Mining processes in dentistry. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, pp. 379–388. ACM (2012)
15. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: a literature review. J. Biomed. Inform. **61**, 224–236 (2016)
16. Rotter, T., Kinsman, L., James, E., Machotta, A., Willis, J., Snow, P., Kugler, J.: The effects of clinical pathways on professional practice, patient outcomes, length of stay, and hospital costs: Cochrane systematic review and meta-analysis. Eval. Health Prof. **35**(1), 3–27 (2012)
17. Song, M., Günther, C.W., van der Aalst, W.M.P.: Trace clustering in process mining. In: Ardagna, D., Mecella, M., Yang, J. (eds.) BPM 2008. LNBIP, vol. 17, pp. 109–120. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00328-8_11
18. Tax, N., Genga, L., Zannone, N.: On the use of hierarchical subtrace mining for efficient local process model mining. In: Proceedings of the 7th International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA 2017), pp. 8–22 (2017)
19. Tax, N., Sidorova, N., Haakma, R., van der Aalst, W.M.P.: Event abstraction for process mining using supervised learning techniques. In: SAI Intelligent Systems Conference 2016, pp. 1–10. IEEE (2016)
20. Tax, N., Sidorova, N., Haakma, R., van der Aalst, W.M.P.: Mining local process models. J. Innov. Digit. Ecosyst. **3**(2), 183–196 (2016)
21. Tax, N., Sidorova, N., van der Aalst, W.M.P., Haakma, R.: Heuristic approaches for generating local process models through log projections. In: IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–8. IEEE (2016)
22. van der Aalst, W.M.P.: Process Mining: Data Science in Action. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
23. van der Aalst, W.M.P., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. Wiley Interdisc. Rev.: Data Min. Knowl. Discov. **2**(2), 182–192 (2012)
24. Verbeek, H.M.W., Buijs, J.C.A.M., Van Dongen, B.F., van der Aalst, W.M.P.: ProM 6: the process mining toolkit. In: Proceedings of the BPM Demonstration Track, vol. 615, pp. 34–39 (2010)
25. Weijters, A.J.M.M., van der Aalst, W.M.P., De Medeiros, A.K.A.: Process mining with the heuristics miner-algorithm. Technical Report WP, Technische Universiteit Eindhoven, vol. 166, pp. 1–34 (2006)

A New Function for Ensemble Pruning

Souad Taleb Zouggar[1(✉)] and Abdelkader Adla[2]

[1] Department of Economics, University of Oran 2, Oran, Algeria
Souad.taleb@gmail.com
[2] Department of Computer Science, University of Oran 1, Oran, Algeria
adla.abdelkader@univ-oran.dz

Abstract. We propose in this work a new function named Diversity and Accuracy for Pruning Ensembles (DAPE) which takes into account both accuracy and diversity to prune an ensemble of homogenous classifiers. A comparative study with a diversity based method and experimental results on several datasets show the effectiveness of the proposed method.

Keywords: Data mining · Classification · Decision trees · Ensemble methods Homogenous ensembles · Bagging · Pruning · Ensemble selection Hill climbing

1 Introduction

Ensemble methods have been proposed to reduce the variance of individual classifiers and consists of two phases: (1) the models generation step in which models can be added without risk of overfitting; these models can be either homogenous or heterogeneous, in the homogenous case models are obtained from the same learning algorithm by varying parameters or using data resampling [6], manipulating input and/or output attributes [1, 7, 12, 17, 29]; (2) the combination step in which the models are aggregated by voting or weighted voting.

However, these methods are known to generate a large number of models which can lead to large storage space and considerable time for classification and prediction. Ensemble selection allows the size reduction of an ensemble consisting of predictive models using different measures based on the diversity and/or the accuracy of the models, this selection allows obtaining smaller ensembles with higher accuracies comparing to the initial ensembles.

Several selection methods have been proposed in the literature; these methods are essentially based on an evaluation function [3, 9, 24] that determines if a model M contributes positively to boost the performances of the whole ensemble. The evaluation is made based on two important properties of ensemble which are diversity and accuracy. Diversity qualifies for a set of classifiers their ability to agree in greater number on good class predictions, and to disagree on classification errors. The diversity and accuracy properties are considered separately [9, 13, 24] or together [4, 11].

In this paper, we propose a novel evaluation function named Diversity and Accuracy for Pruning Ensembles (DAPE) based on both diversity and accuracy, the method is

F. Dargam et al. (Eds.): ICDSST 2018, LNBIP 313, pp. 181–190, 2018.
https://doi.org/10.1007/978-3-319-90315-6_15

applied on homogenous ensembles composed of C4.5 decision trees [26]. This method is based on a DHCEP (Directed Hill Climbing Ensemble Pruning) strategy with a multi-objective function to evaluate the relevance of an ensemble of trees. The function, used in a Hill Climbing process in Forward Selection (FS), allows selection of ensembles with the best compromise between maximum diversity and minimum error rate.

A comparative study of the proposed measure and the UWA measure [25] is carried out on datasets from the UCI Repository [2].

The paper is structured as follows: In Sect. 2 we present the bagging used for homogenous ensemble generation, the combination using weighted and unweighted voting and finally the hill climbing method used for the search of a solution. Section 3 gives a review on recent works in the same domain. Section 4 introduces the proposed method in details. The penultimate section contains a comparative study and experimental results on multiple datasets. Finally we conclude and give some future works.

2 Background

In this section, we highlight the basic elements used in this paper, namely, the bagging method used for the generation of the initial ensemble, aggregation by unweighted and weighted vote.

2.1 Diversification by Bagging

Bagging **B**ootstrap **Agg**regat**ing**, a resampling method introduced by Breiman in 1996 [6]. Given a learning sample Ω_L and a learning method which generates a predictor $\hat{h}(.,\Omega_L)$ using Ω_L. The principle of bagging is to draw several bootstrap samples $\left(\Omega_L^{\theta 1}, \ldots, \Omega_L^{\theta q}\right)$ and generate for each one a collection of predictors $(\hat{h}(., \Omega_L^{\theta 1}), \ldots, \hat{h}(., \Omega_L^{\theta q}))$ using the base learning method for finally aggregating them.

A bootstrap sample Ω_a^l is obtained by randomly drawing n observations in the starting sample Ω_L. Each observation have the probability of *1/n* of being shot; $|\Omega_L| = n$, the random variable θ_l represents the random drawing.

Initially, Bagging was introduced with a decision tree as basic rule. But the schema is general and can be apply to other basic rules. Bagging transforms an unstable algorithm into a rule with very good properties (consistency and optimal speed of convergence) [5] (Fig. 1).

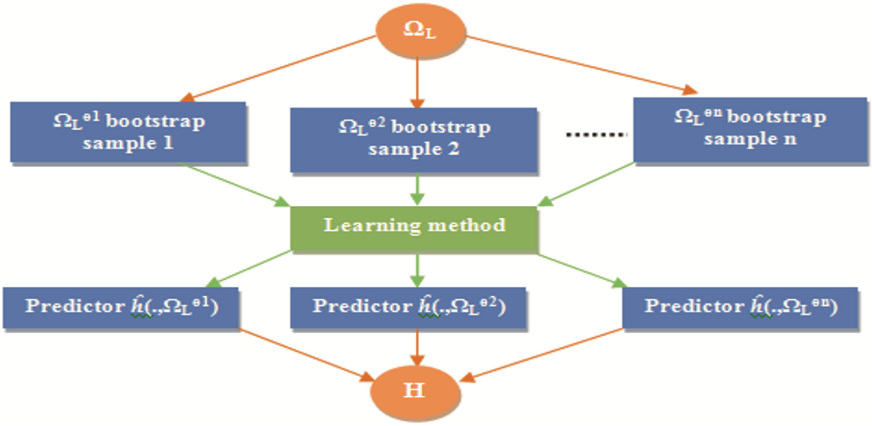


Fig. 1. Representative diagram of bagging.

2.2 Aggregation (Unweighted Vote, Weighted Vote)

The unweighted and weighted voting are the most used methods for combining (aggregating) whether homogenous or heterogeneous models. In ensemble methods each model, for an instance, gives a class value, a probability, and the class with most votes, highest average probability is assigned to the instance by the ensemble.

In weighted vote, the classification models are associated with weights assigned relatively to their classification accuracy. Formally this can be written: [25] Let x be an instance and $m_{i,i=1..k}$ a set of models that output a probability distribution $m_i(x, c_j)$ for each class $c_{j,j=1...n}$. The output of the (weighted) voting method y(x) for instance x is given by the following mathematical expression:

$$y(x) = argmax_{c_j} \sum_{i=1}^{k} w_i m_i(x, c_j) \tag{1}$$

Where w_i is the weight of model i. In the simple case of voting (unweighted), the weights are all equal to one, that is, $w_{i,i=1...k} = 1$.

2.3 The Hill Climbing

Hill climbing is an optimization technique belonging to the family of local search. The algorithm starts with any solution to a problem, and then tries iteratively to find a better solution by changing one element of the solution. If the change produces a better solution (maximize or minimize the evaluation function used for the course), an incremental change is made to the new solution. The process is repeated until no improvements can be found (the function reached the maximum or the minimum).

Hill climbing attempts to maximize (or minimize) a target function f(X) where X is a vector of continuous and/or discrete values. Each iteration, hill climbing will adjust a single element in X and determine if the change improves the value of f(X). Any change improving the function *f*(X) is accepted, the process continues until no amelioration of the function can be found.

For ensemble selection, DHCEP (Directed Hill Climbing Ensemble Pruning) is used, in this case the vector X is composed of classifiers or predictors. The path can be realized either in backward elimination or in forward selection, in the first case the whole ensemble is considered as a solution and then repeatedly elements not improving the evaluation function are eliminated one by one, in the second case we initialize with an element randomly and we add the elements that improve the evaluation function one by one. The elements to be added or removed are part of the neighborhood of the current solution (Fig. 2).

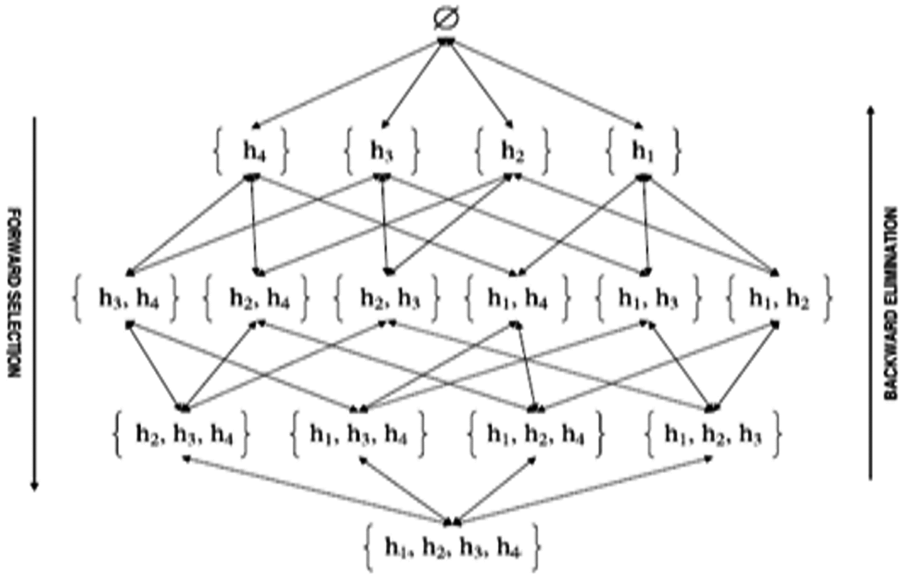


Fig. 2. Hill climbing search for selection in an ensemble composed of four classifiers [25].

3 Related Work

Several methods have been proposed to reduce the size of a set of classifiers [14, 16, 18, 19, 21, 23, 27]. The methods differ from each other by the adopted research directions, the different evaluation measures or the evaluation ensembles used.

A first type of approaches uses performance measures. Fan et al. [13] propose a profit-based evaluation function and propose dynamic scheduling to accelerate the prediction process. For the reduction of the ensemble size, the total benefit is used as selection criterion in conjunction with a greedy search algorithm with and without back fitting. The path begins with the model with the greatest benefit. A set of instances x is considered, each instance (x) can be positive or negative, $B(x)$ denotes the benefit of predicting x as positive and the total benefit $BT = \sum_x B(x)$, the authors choose the sub ensembles which maximize the total benefit. Caruana et al. [9] use several performance metrics and a hill climbing strategy for building ensembles of models from libraries of thousands of models. Model libraries are generated using different learning algorithms. The Forward Stepwise selection is used to add to the ensemble the models that maximize its performance.

The second type of approaches uses diversity-based measures [3, 20, 24]. Partalas et al. [25] propose a measure of diversity that considers all the cases that may exist when adding a model h_t to an ensemble. The measure Uncertainty Weighted Accuracy (UWA) considers four cases when adding and using justified weights to distinguish favor cases from others.

$$UWA_{Eval}(h_k, Sub) = \sum_{i=1}^{|Eval|} \big(\alpha * I(y_i = h_k(x_i) ET y_i \neq Sub(x_i)) - \beta * I(y_i \neq h_k(x_i) ET y_i = Sub(x_i)) + \beta * I(y_i = h_k(x_i) ET y_i = Sub(x_i)) - \alpha * I(y_i \neq h_k(x_i) ET y_i \neq Sub(x_i))\big) \quad (2)$$

The parameters α, β represent respectively the number of models in the sub-set Sub correctly classifying the instance (x_i, y_i) and the number of models incorrectly classifying the same instance.

Li et al. [22] theoretically deals with the effect of diversity on voting generalization performance using Probably Approximately Correct (PAC) learning. It is revealed that diversity is closely related to the space complexity hypothesis, and strengthening it can be achieved by applying regularization to ensemble methods. Based on this analysis, the authors apply an explicit regularization of the diversity for the selection of ensembles.

Zhou et al. [30] propose a new algorithm based on frequent item learning that links data and the simplified ensemble to a transactional database whose transactions are instances and items are classifiers. A Boolean classification matrix is used for each model of the pruned ensemble. Using this matrix, several candidate ensembles are obtained by iterative and incremental extraction of basic classifiers with the best performances.

Bhatnagar et al. [4] perform ensemble selection using a performance-based and diversity-based function that considers the individual performance of classifiers as well as the diversity between pairs of classifiers. A bottom-up search is performed to generate the sub ensembles by adding various pairs of classifiers with high performance.

Cavalcanti et al. [10] combine in pairs different matrices of diversity using a genetic algorithm. The combined diversity matrix is then used to group similar (not very diverse) models; they must not belong to the same ensemble. To generate candidate ensembles, the combined diversity matrix is transformed into one or more graphs and then a graph coloring technique is applied.

Guo et al. [15] propose a new metric using the margin (instances) and the diversity (of classifiers) to explicitly evaluate the importance of individual classifiers. By adding the models to the ensemble in decreasing order of the metric, the user can choose the first T models to form a sub-ensemble.

Qun et al. [11] emphasize the utility of optimizing predictive performance together with diversity, which are two indispensable and inseparable parameters for ensemble selection. There have been three measures proposed to simplify ensembles using a greedy algorithm: (1) The first measure simultaneously considers the difference (diversity) between the current subset and the candidate classifier and the performance of each one; (2) The second allows evaluating the diversity within the ensemble and; (3) the last measure reinforces the concern about the accuracy of the resulting sub-ensemble. Experimental results confirm the interest of the three measures which is illustrated by the improvement of performances.

4 The Proposed Method

The set of data Ω is divided into two sub samples Ω_L (generally 80% of Ω) for learning and pruning and Ω_T (generally 20% of Ω) for testing. A bagging ensemble BE of t C4.5 trees is constructed, $\mathrm{BE} = \{\mathrm{T}_1, \ldots, \mathrm{T}_\mathrm{i}, \ldots, \mathrm{T}_\mathrm{t}\}$, using Ω_L with $|\Omega_L| = n$. Each tree T_i is represented by a vector $(x_{1i}, x_{2i}, \ldots, x_{ji}, \ldots, x_{ni})^T$. We have the following notations:

- x_{ij}: Result of classification of the individual i by the tree j, $x_{ij} = 1$ if the individual i is misclassified by the tree T_j and $x_{ij} = 0$ otherwise,
- x_{i+}: The total number of errors committed for the individual I:

$$x_{i+} = \sum_{j=1}^{t} x_{ij} \tag{3}$$

- X: The total number of errors committed by the set,

$$X = \sum_{i=1}^{n} \sum_{j=1}^{t} x_{ij} \tag{4}$$

- (Θ_i, x_{i+}): The relative distribution of the error frequencies associated with the different cases,

$$\theta_i = \frac{x_{i+}}{X}, \ i = 1, n \tag{5}$$

- x_{+j}: The number of errors committed by the classifier Tj over all the individuals,

$$x_{+j} = \sum_{i=1}^{n} x_{ij} \tag{6}$$

- e_j: The error rate associated with the tree T_j,

$$e_j = \frac{x_{+j}}{n} \tag{7}$$

The evaluation function to optimize noted S connects diversity Θ_i and the error rate e_j (α is a parameter for which values are chosen empirically):

$$S = \sum_{i=1,n} \theta_i^2 + \alpha \sum_{j=1,t} e_j^2 \tag{8}$$

The algorithm below presents the proposed method DAPE in a pseudocode:

```
Algorithm DAPE;

Input
B = {T1,...,Tt};
ΩT: selection set;
Neighborhood(Ψj): Function that returns the subsets of
models obtained from Ψj by adding a classifier (tree);
Output
Sub ensemble Ψ0 of B;
Begin
Initialize(Ψ0);
1. Calculate S(Ψ0,ΩT);
   If ∃ Ψj such as S(Ψj,ΩT) < S(Ψ0,ΩT) where Ψj ∈
   Neighborhood(Ψ0)Then Ψ0=argminΨj(S(Ψj,ΩT));
   Goto 1;
End.
```

5 Experiments

The experiments consist in building homogeneous sets by sampling the starting sample and using the C4.5 decision tree generation algorithm as a basic rule [26]. The Weka platform [28] is used as a source for the C4.5 learning algorithm and validation. We consider 10 data sets from the UCI Repository [2] which are described in Table 1:

Table 1. Description of datasets.

	Number of instances	Number of descriptors	Class modalities
Breast w	699	9	2
Tic tac toe	958	9	2
Dermatology	366	34	6
ecoli	336	7	8
kr-vs-kp	3196	36	2
glass	214	9	6
heart-h	294	13	5
hepatitis	155	19	2
ionosphere	351	34	2
lymph	148	18	4

The proposed method, DAPE, is compared to the ensemble pruning method based on diversity UWA [25] detailed in literature and whose source code is available at http://mlkd.csd.auth.gr/ensemblepruning.html. For two methods, the unweighted majority vote is used for the combination of models and the performance calculation.

The methods use a forward selection strategy in a hill climbing scheme. The stop criterion for UWA is the performance on the evaluation sample which generates subsets of reduced sizes compared with the usual stop criteria defined as a fixed number of models [24]. In our case we use the same function for both the path and the stop.

We use two criteria to compare the pruning methods; the performance and the size of the subsets obtained, ALL designs the ensemble composed of all the models (Tables 2 and 3).

Table 2. Comparativestudyof DAPE and UWA based on accuracy of obtained sub ensembles.

	DAPE	UWA	ALL
Breast w	**0.97459**	0.95032	0.95282
Tic tac toe	0.94205	**0.95338**	0.9333
Dermatology	0.94861	**0.96712**	0.91369
ecoli	**0.84912**	0.84326	0.83429
kr-vs-kp	0.98622	0.99433	**0.9962**
glass	**0.76882**	0.7619	0.76509
heart-h	**0.80951**	0.79652	0.79658
hepatitis	0.81613	0.80965	**0.81932**
ionosphere	**0.9235**	0.90856	0.93425
lymph	**0.8331**	0.81376	0.76997
Average success rates	**0.885165**	0.87988	0.871551

Table 3. Comparative study of DAPE and UWA based on the size of obtained sub ensembles.

	DAPE	UWA
Breast w	**9.8**	11.7
cmc	**18**	48.7
Dermatology	10	**8.7**
ecoli	12.1	**11.5**
kr-vs-kp	7.5	**6.2**
glass	**13**	15.9
heart-h	**13.2**	15.7
hepatitis	**10.1**	13
ionosphere	10.6	**6.7**
lymph	10.9	**10.6**
Average sizes	**11.31**	12.31

For 6 out of 10 benchmarks, DAPE shows better performances compared to the UWA which scores only 2 better performances, the whole ensemble gives better performances in only 2 cases.

For the average success rate on all datasets, DAPE is ranked first with a rate of 0.885165, exceeding UWA with a rate of 0.5%, and the whole ensemble with 1%.

All the methods DAPE, DEAS and UWA allow obtaining subsets of reduced sizes for 5 cases, 50% of the cases. For the average size on the 10 datasets, DAPE is ranked first with an average size of 11.31.

6 Conclusion and Future Work

In this paper we presented a new evaluation function combining performance and diversity for selection in a homogeneous ensemble used in a search process based on climbing hill. The method was evaluated on several benchmarks and compared to the method UWA based only on diversity, the results show the superiority of the proposed method.

We used the DAPE method for the selection in heterogeneous ensembles; where the classifiers are not generated from the same learning algorithm and for the selection in random forest ensembles, knowing that a random forest ensemble improves the performance of a bagging [8].

We also propose to study the possibility of using another search strategy for the selection in order to reduce the search time of the ensembles because the hill climbing strategy requires a non-negligible time when the number of models becomes important.

The last point consists in finding a value for the parameter α for which we have noticed during empirical research that for which an appropriate value could significantly improve the results as observed in our experiments.

References

1. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Comput. **9**, 1545–1588 (1997)
2. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007). http://www.ics.uci.edu/mlearn/MLRepository.html
3. Banfield, R.E., Hall, L.O., Bowyer, K.W., Kegelmeyer, W.P.: Ensemble diversity measures and their application to thinning. Inf. Fusion **6**(1), 49–62 (2005)
4. Bhatnagar, V., Bhardwaj, M., Sharma, S., Haroon, S.: Accuracy-diversity based pruning of classifier ensembles. Prog. Artif. Intell. **2**(2–3), 97–111 (2014)
5. Biau, G., Cérou, F., Guyader, A.: On the rate of convergence of the bagged nearest neighbor estimate. J. Mach. Learn. Res. **11**, 687–712 (2010)
6. Breiman, L.: Bagging predictors. Mach. Learn. **26**(2), 123–140 (1996)
7. Breiman, L.: Randomizing outputs to increase prediction accuracy. Mach. Learn. **40**, 229–242 (2000)
8. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
9. Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A.: Ensemble selection from libraries of models. In: Proceedings of the 21st International Conference on Machine Learning (2004)
10. Cavalcanti, G.D.C., Oliveira, L.S., Moura, T.J.M., Carvalho, G.V.: Combining diversity measures for ensemble pruning. Pattern Recognit. Lett. **74**, 38–45 (2016). ISSN 0167-8655
11. Qun, D., Ye, R., Liu, Z.: Considering diversity and accuracy simultaneously for ensemble pruning. Appl. Soft Comput. **58**, 75–91 (2017). ISSN 1568-4946
12. Dietterich, T.G., Kong, E.B.: Machine learning bias, statistical bias, and statistical variance of decision tree algorithms. Technical report, Dept of Computer Science, Oregon State University, Covallis, Oregon (1995)
13. Fan, W., Chu, F., Wang, H., Yu, P.S.: Pruning and dynamic scheduling of cost-sensitive ensembles. In: Eighteenth National Conference on Artificial Intelligence, pp. 146–151. American Association for Artificial Intelligence (2002)
14. Fu, Q., Hu, S.X., Zhao, S.Y.: Clustering-based selective neural network ensemble. J. Zhejiang Univ. Sci. A **6**(5), 387–392 (2005)

15. Guo, H., Liu, H., Li, R., Wu, C., Guo, Y., Xu, M.: Margin & diversity based ordering ensemble pruning. Neurocomputing **275**, 237–246 (2017). ISSN 0925-2312
16. Hernández-Lobato, D., Martínez-Munoz, G.: A statistical instance-based pruning in ensembles of independent classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 364–369 (2009)
17. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
18. Margineantu, D., Dietterich, T.G.: Pruning adaptive boosting. In: Proceedings of the 14th International Conference on Machine Learning, pp. 211–218. Morgan Kaufmann, San Francisco (1997)
19. Markatopoulou, F., Tsoumakas, G., Vlahavas, I.: Instance-based ensemble pruning via multi-label classification. In: ICTAI 2010 (2010)
20. Martínez-Muñoz, G., Suárez, A.: Aggregation ordering in bagging. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Applications, pp. 258–263. Acta Press (2004)
21. Martínez-Muñoz, G., Hernández-Lobato, D., Suárez, A.: Selection of decision stumps in bagging ensembles. In: de Sá, J.M., Alexandre, L.A., Duch, W., Mandic, D. (eds.) ICANN 2007. LNCS, vol. 4668, pp. 319–328. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74690-4_33
22. Li, N., Yu, Y., Zhou, Z.-H.: Diversity regularized ensemble pruning. In: Flach, P.A., De Bie, T., Cristianini, N. (eds.) ECML PKDD 2012. LNCS (LNAI), vol. 7523, pp. 330–345. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33460-3_27
23. Partalas, I., Tsoumakas, G., Katakis, I., Vlahavas, I.: Ensemble pruning using reinforcement learning. In: Antoniou, G., Potamias, G., Spyropoulos, C., Plexousakis, D. (eds.) SETN 2006. LNCS (LNAI), vol. 3955, pp. 301–310. Springer, Heidelberg (2006). https://doi.org/10.1007/11752912_31
24. Partalas, I., Tsoumakas, G., Vlahavas, I.: Focused ensemble selection: a diversity-based method for greedy ensemble selection. In: Ghallab, M., Spyropoulos, C.D., Fakotakis, N., Avouris, N.M. (eds.) ECAI 2008 - 18th European Conference on Artificial Intelligence. Proceedings of the Frontiers in Artificial Intelligence and Applications, Patras, Greece, 21–25 July 2008, vol. 178, pp. 117–121. IOS Press (2008)
25. Partalas, I., Tsoumakas, G., Vlahavas, I.: An ensemble uncertainty aware measure for directed hill climbing ensemble pruning. Mach. Learn. **81**, 257–282 (2010)
26. Quinlan, J.R.: C4.5 Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
27. Soto, V., Martínez-Muñoz, G., Hernández-Lobato, D., Suárez, A.: A double pruning algorithm for classification ensembles. In: El Gayar, N., Kittler, J., Roli, F. (eds.) MCS 2010. LNCS, vol. 5997, pp. 104–113. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12127-2_11
28. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, Los Altos (2005)
29. Zheng, Z., Webb, G.I.: Stochastic attribute selection committees. Technical report (TR C98/08), School of Computing and Mathematics, Deakin University, Australia (1998)
30. Zhou, H., Zhao, X., Wang, X.: An effective ensemble pruning algorithm based on frequent patterns. Knowl.-Based Syst. **56**, 79–85 (2014). ISSN 0950-7051

Author Index