Università degli Studi di Salerno

Dottorato di Ricerca in Informatica e Ingegneria dell'Informazione
Ciclo 30 – a.a 2016/2017

TESI DI DOTTORATO / PH.D. THESIS

Cognitive Models and Computational Approaches for improving Situation Awareness Systems

GIUSEPPE D'ANIELLO

SUPERVISOR: **PROF. MATTEO GAETA**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Dipartimento di Ingegneria dell'Informazione ed Elettrica
e Matematica Applicata
Dipartimento di Informatica

To my family and
to my lovely Anna Teresa.

Behind every beautiful thing,
there's some kind of pain.

- Bob Dylan, *Not Dark Yet* -

Abstract

The world of *Internet of Things* is pervaded by complex environments with smart services available every time and everywhere. In such a context, a serious open issue is the capability of information systems to support adaptive and collaborative decision processes in perceiving and elaborating huge amounts of data. This requires the design and realization of novel socio-technical systems based on the "human-in-the-loop" paradigm. The presence of both humans and software in such systems demands for adequate levels of Situation Awareness (SA). To achieve and maintain proper levels of SA is a daunting task due to the intrinsic technical characteristics of systems and the limitations of human cognitive mechanisms. In the scientific literature, such issues hindering the SA formation process are defined as SA demons.

The objective of this research is to contribute to the resolution of the SA demons by means of the identification of information processing paradigms for an original support to the SA and the definition of new theoretical and practical approaches based on cognitive models and computational techniques.

The research work starts with an in-depth analysis and some preliminary verifications of methods, techniques, and systems of SA. A major outcome of this analysis is that there is only a limited use of the Granular Computing paradigm (GrC) in the SA field, despite the fact that SA and GrC share many concepts and principles. The research work continues with the definition of contributions and original results for the resolution of significant SA demons, exploiting some of the approaches identified in the analysis phase (i.e., ontologies, data mining, and GrC). The first contri-

bution addresses the issues related to the bad perception of data by users. We propose a semantic approach for the quality-aware sensor data management which uses a data imputation technique based on association rule mining. The second contribution proposes an original ontological approach to situation management, namely the Adaptive Goal-driven Situation Management. The approach uses the ontological modeling of goals and situations and a mechanism that suggests the most relevant goals to the users at a given moment. Lastly, the adoption of the GrC paradigm allows the definition of a novel model for representing and reasoning on situations based on a set theoretical framework. This model has been instantiated using the rough sets theory. The proposed approaches and models have been implemented in prototypical systems. Their capabilities in improving SA in real applications have been evaluated with typical methodologies used for SA systems.

Acknowledgements

Almost there, tomorrow I will submit the thesis, and I take a moment to stop and to look back at these years. I have to recognize that many people have contributed and made possible to achieve this goal, and I would like to thank all of them.

It is really difficult to express in few words all my gratitude to who more than anyone else has made all this possible. To my supervisor, mentor and friend, Prof. Matteo Gaeta, I am grateful for his guidance, his motivation, his support and his teachings. You convinced me to start this journey (and today I thank you) and together we reached important goals and results. Thanks for all the opportunities you gave me in these years.

A special thank goes to Prof. Vincenzo Loia, one of the greatest scholar that I have ever had the fortune to meet. Thanks for having introduced me to the interesting topic of Situation Awareness, for having guided me during the research work and for all the opportunities you gave me.

I would also thank all the members of the Ph.D. Program board, in particular the Ph.D. Program Director, Prof. Pasquale Chiacchio, and the Director of the Department of Information and Electrical Engineering and Applied Mathematics, Prof. Mario Vento, for their support and the constant stimuli. Many thanks to the evaluators of this thesis for their precious work.

During these years, I learned how much the research collaborations are fundamental to achieve great results. For this, I thank all the people with whom I have collaborated. I thank Prof. Francesco Orciuoli for all the time he dedicated to me. Much of this work has been made with you, and if today I am able to write a paper the merit is also yours. Thanks to Dr. Angelo Gaeta for all the interesting discussions and all the work we did together.

I thank all the fellows of the Research Consortium on Agent Systems (CORISA), in particular the director Prof. Massimo de Falco for having given me the opportunity to work on such challenging and interesting research projects that contribute to the development of this thesis. I also thank Eng. Mario Lepore: a sincere and always helpful colleague, a true friend, a skillful professional, whose suggestions and help have been fundamental for this work.

I have also had the opportunity to collaborate with many exceptional people all around the world, and to them I wish to extend my warmest thanks. In particular, I would like to thank Prof. Tzung-Pei Hong, a true model for all of us young researchers, whose suggestions and teachings have been inspirational for my work. Special thanks goes to Prof. Marek Reformat. Besides all your great recommendations and suggestions on my research work and on the academic life, I wish to thank you so much for all your encouragements and the long conversations we made in these years (so thank you too, Skype!).

The journey towards the Ph.D. is really long and winding, with many ups and downs. Having been able to count on good friends made this work more pleasant. Listing all of you here is difficult. Hoping that the others do not resent, I have to thank a special couple that always believed in me, made me laugh, and supported me: thank you Alfonso and Chiara (and even if you don't know me yet, thank you baby girl, we're waiting for you!).

Lastly, I would like to thank my parents. *Grazie mamma e papà per il vostro supporto, per avermi permesso di studiare liberamente in tutti questi anni, per il vostro affetto e per il vostro duro lavoro che è di esempio per me. Grazie.* I also wish to thank my big brother Catello, Sira and my lovely boys Luca and Vincenzo.

I have no words to express all of my gratitude to a very special woman, my beloved Anna Teresa for always being there for me. Without you, I would probably have given up long ago. And that is why I am sure together we will realize our dreams and achieve

the most important goals of our life.

A thought goes to who is no longer here physically, but that never left me during these years. Hope you could see all of this, wherever you are.

Giuseppe D'Aniello

List of Publications

Parts of this thesis are based on the following publications. As a coauthor, I was involved actively in the research, planning and writing these papers.

International Journals

I Giuseppe D'Aniello, Matteo Gaeta, Tzung-Pei Hong: Effective Quality-aware Sensor Data Management, IEEE Transactions on Emerging Topics in Computational Intelligence, vol. PP, no. 99, pp. 1-13, 2018, ISSN 2471-285X, DOI: 10.1109/TETCI.2017.2782800 .

II Giuseppe D'Aniello, Matteo Gaeta, Francesca Loia, Marek Reformat, Daniele Toti: An Environment for Collective Perception based on Fuzzy and Semantic Approaches. Journal of Artificial Intelligence and Soft Computing Research, 2018, Vol. 8, No. 3, pp. 191-210 ISSN 2083-2567, DOI: 10.1515/jaiscr-2018-0013.

III Giuseppe D'Aniello, Matteo Gaeta, Francesco Orciuoli: An Approach based on Semantic Stream Reasoning to support Decision Processes in Smart Cities. Telematics and Informatics 35(1), pp. 68-81, 04/2018, ISSN 0736-5853, DOI: 10.1016/j.tele.2017.09.019.

IV Giuseppe D'Aniello, Angelo Gaeta, Vincenzo Loia, Francesco Orciuoli. A granular computing framework for approximate reasoning in situation awareness. Journal of Granular Computing, 2(3), 11/2016, pp 141-158, ISSN 2364-4966, DOI: 10.1007/s41066-016-0035-0.

V Vincenzo Loia, Giuseppe D'Aniello, Angelo Gaeta, Francesco Orciuoli: Enforcing situation awareness with granular computing: A systematic overview and new perspectives. Journal of Granular Computing, 1(2), 01/2016; ISSN 2364-4966, DOI:10.1007/s41066-015-0005-y

VI Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Stefania Tomasiello: Self-regulated learning with approximate reasoning and situation awareness. Journal of Ambient Intelligence and Humanized Computing 10/2016, ISSN 1868-5137, DOI:10.1007/s12652-016-0423-y

VII Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Mario Lepore, Francesco Orciuoli, Orlando Troisi: A new DSS based on situation awareness for smart commerce environments. Journal of Ambient Intelligence and Humanized Computing 07/2015; 7(1):1-15., ISSN 1868-5137, DOI:10.1007/s12652-015-0300-0

VIII Giuseppe D'Aniello, Vincenzo Loia, Francesco Orciuoli. A Multi-Agent Fuzzy Consensus Model in a Situation Awareness Framework. Applied Soft Computing 02/2015; 30, pp. 430-440, ISSN 1568-4946 DOI:10.1016/j.asoc.2015.01.061

International Conference Proceedings

I Giuseppe D'Aniello, Angelo Gaeta, Vincenzo Loia, Francesco Orciuoli: A model based on rough sets for situation comprehension and projection. 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management, CogSIMA 2017, art. no. 7929578, DOI:10.1109/COGSIMA.2017.7929578 **(Best Paper Award)**

II Giuseppe D'Aniello, Matteo Gaeta, Marek Z. Reformat: Collective Perception in Smart Tourism Destinations with Rough Sets. 2017 3rd IEEE International Conference on Cybernetics (CYBCONF); 06/2017, DOI:10.1109/CYBConf.2017.7985765

III Giuseppe D'Aniello, Vincenzo Loia, Francesco Orciuoli: Adaptive Goal Selection for improving Situation Awareness: the Fleet Management case study. Procedia Computer Science (8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017) 12/2017; 109:529-536., DOI:10.1016/j.procs.2017.05.332

IV Giuseppe D'Aniello, Massimo de Falco, Matteo Gaeta, Francesca Loia: Analytical Marketing with Collective Perception. 1st International Conference on Advances in Business, Management and Law (2017), Dubai, U.A.E.; 11/2017

V Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Vincenzo Loia, Marek Z. Reformat: Application of Granular Computing and Three-way decisions to Analysis of Competing Hypotheses. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016), Budapest, Hungary; 10/2016, DOI:10.1109/SMC.2016.7844475.

VI Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Vincenzo Loia, Marek Z. Reformat: Collective Awareness in Smart City with Fuzzy Cognitive Maps and Fuzzy sets. 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, Canada; 07/2016, DOI:10.1109/FUZZ-IEEE.2016.7737875

VII Giuseppe D'Aniello, Angelo Gaeta, Vincenzo Loia, Francesco Orciuoli: Integrating GSO and SAW Ontologies to enable Situation Awareness in Green Fleet Management. 2016 IEEE International Multi-Disciplinary Conference on Cognitive

Methods in Situation Awareness and Decision Support (CogSIMA), San Diego, USA; 03/2016, pp. 138-144, DOI:10.1109/COGSIMA.2016.7497801

VIII Giuseppe D'Aniello, Angelo Gaeta, Francesco Orciuoli: Artificial bees for improving resilience in a sensor middleware for situation awareness. The 2015 Conference on Technologies and Applications of Artificial Intelligence (TAAI 2015), Tainan, Taiwan; 11/2015, pp. 300-307. DOI:10.1109/TAAI.2015.7407104

IX Giuseppe D'Aniello, Vincenzo Loia, Francesco Orciuoli: Employing Fuzzy Consensus for Assessing Reliability of Sensor Data in Situation Awareness Frameworks. 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC2015, Hong Kong; 10/2015, pp. 2591-2596 DOI:10.1109/SMC.2015.453

X Giuseppe D'Aniello, Matteo Gaeta, Vincenzo Loia, Francesco Orciuoli, Demetrios G. Sampson: Situation Awareness Enabling Decision Support in Seamless Learning. 2015 International Conference on Intelligent Networking and Collaborative Systems (IEEE INCoS 2015), Taipei; 09/2015, pp. 440-445, DOI:10.1109/INCoS.2015.59

XI Giuseppe D'Aniello, Matteo Gaeta, Antonio Granito, Vincenzo Loia, Francesco Orciuoli: Sustaining Self-Regulation Processes in Seamless Learning Scenarios by Situation Awareness. 2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Orlando, Florida. USA; 03/2015, pp. 101-105, DOI:10.1109/COGSIMA.2015.7108182

Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Definition
AGSM	Adaptive Goal-driven Situation Management
ANN	Artificial Neural Network
CEP	Complex Event Processing
CST	Context Space Theory
DSS	Decision Support System
DST	Dempster-Shafer Theory
FCA	Formal Concept Analysis
FCM	Fuzzy Cognitive Map
GDTA	Goal-Directed Task Analysis
GFMS	Green Fleet Management System
GrC	Granular Computing
GS	Granular Structure
HMM	Hidden Markov Model
IG	Information Granularity
IoT	Internet of Things
IT	Information Table
KPI	Key Performance Indicator
LTM	Long Term Memory
MLN	Markov Logic Network
NS	Neighborhood System

Acronym	Definition
OWL	Web Ontology Language
POMPDP	Partially Observable Markov Decision Process
QoI	Quality of Information
QoS	Quality of Service
RAR	Robust Association Rule
RDF	Resource Description Framework
RMSE	Root Mean Square Error
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SLA	Service Level Agreement
SN	Sensor Network
SOM	Self-Organizing Map
SPARQL	SPARQL Protocol and RDF Query Language
SSNO	Semantic Sensor Network Ontology
ST	Situation Theory
STO	Situation Theory Ontology
SVD	Single Value Decomposition
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
VS	Virtual Sensor
WM	Working Memory

Chapter 1

Introduction

> *"Can we get serious now?*
> *We've all heard about the computer simulations, and now we are watching actual sims, but I can't quite believe you still have not taken into account the human factor."*
>
> — Chesley Sullenberger, *Sully*

We live in the *knowledge society* – this means that our economy and our lives are based on the processing of information and on the exploitation of the knowledge contained in it. In such a context, owning and mastering the right information represents a tremendous competitive advantage. Nevertheless, the main problem is not the lack of available information. Rather, the issue is exactly the opposite, namely the overload of data. There are a few main causes that have led to huge and growing volumes of data. Some causes include the growth of the *Internet of Things* paradigm, with the always more pervasive availability of new and more intelligent sensors, the *Big Data* phenomenon where massive data are generated at an extremely rapid rate, and the increasing storage capacity with the availability of cheaper devices. However, regardless of the human ability to produce and store the data effectively, the capability to analyze and exploit such data for various activities – such as performing tasks, executing process and making decisions – has been quickly outpaced.

In spite of this vastness of data, many individuals may find themselves nowadays to be less informed than ever before [9]. This is mainly due to the gap between the huge quantity of generated information and the human ability to identify the pieces of data needed to perform tasks and to make decisions. It is necessary to deal with such challenging gap to be able to make decisions continuously – both in professional and private spheres – often by interacting with systems and tools that try to support these tasks. However, the realization of increasingly sophisticated and complex systems capable of supporting and automatizing decisions is not sufficient. Due to the complex and dynamic nature of many real-world problems, it is not always possible to address the problems of data processing, analysis and decision making, by using technology exclusively. To date, many cases show the necessity for human involvement in the loop of monitoring and control of such systems, since only the cognitive capabilities of humans are able to deal with such dynamism and complexity. Consequently, this requires the design and realization of novel socio-technical systems in which humans and machines coexist and work synergistically to achieve a common objective. Such systems are based on the so-called "human-in-the-loop" paradigm, and when they are correctly designed, they are a valid ally for reducing this information gap. Regrettably, the co-existence of humans factors and system characteristics introduce novel and daunting challenges for the decision-making processes. Individuals have different execution times greater than that of machines; they suffer from data and work overload, and there is a greater tendency for them to make errors due to having wrong mental models, poor concentrations and to over-focus on wrong information. On the other hand, the high level of automation of novel systems has exacerbated the problem rather than contributing to solve it. This level of sophistication makes it even more troublesome to support the interaction with the human users as a lot of stress has been placed on them.

This situation led to numerous accidents in the last 20 years in large scale socio-technical systems and control rooms with disastrous consequences. Often, the causes of such accidents are simplistically attributed either to technical errors and faults or to generic human errors. A more in-depth analysis shows that such errors are mainly attributed to the interactions between human users and the systems, specifically in the lack of awareness of the human users regarding the actual current situations and the knowledge about the actual state of the systems. This is the case with many famous tragic accidents of recent years, such as the Air France Flight 447 which crashed into the Atlantic Ocean on June 1, 2009 due to an incorrect reaction of the crew to technical failures [10]. Another instance is the disaster at the nuclear power plant of Chernobyl, in which the investigation was concluded by attributing the root cause to the "so-called human element"[11]. There are also many maritime disasters, such as the case of MS Estonia in 1994 in which 852 persons died, which happened also due to wrong actions by the crew [12]. Many other accidents in different domains can be attributable to human factors; specifically, the investigations and the opinions of the scientific community [13, 14, 15] demonstrate that the root cause of such accidents and errors can be found in the lack of situation awareness (SA). Situation awareness can be considered as the cognitive capability of an operator to understand what is happening in the environment. It can be described as the level of awareness that a person has with respect to the task currently performed. SA is often considered as the most important prerequisite for making correct decisions. Endsley defines the concept as "the perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future"[1]. Many works have shown that the lack of SA is among the root cause of human errors; such a lack has accounted for at least 85% of errors in the aviation domain [14], 80% in the medical domain [16], 70% in the maritime domain [17, 18].

1.1 Research Challenges and Objectives

Having identified SA as the root cause of many human errors and accidents, the next step is to explore the reasons behind the challenges of reaching and maintaining an adequate awareness regarding current or on-going situations as well as understanding these situations and reacting accordingly. It is also important to determine the factors that undermine the process of SA assessment (i.e., the process of gaining and maintaining high level of SA).

The research community has worked long on this topic, particularly in identifying the main causes and factors that human users struggle with in achieving good SA. Endsley, after having identified a set of typical errors in SA [14], has synthesized and defined the eight common causes for a lack of SA, which has been called the Demons of Situation Awareness (SA demons). These eight causes are attentional tunneling, data overload, complexity, memory trap, workload and stressors, wrong mental models, misplaced salience, and out-of-the-loop [15]. They occur as a result of individual factors as well as system or environmental factors. Specifically, some of the cognitive factors that may give rise to the SA demons include the limited working memory, the presence of expectations and preconceptions, the mental models, and the automaticity in performing some tasks. The main system and environmental characteristics that make it difficult to solve the SA demons include system complexity, the interface design, the stressors and the workload, and automation.

A proper design of the system and its interface is fundamental for addressing the SA demons, as such a design would support the process of SA assessment of users. Endsley has proposed a set of best practices and principles in her book [15], and these practices help in the definition of systems capable of improving the level of SA of the users by softening the issues with SA demons. However, this alone may not be sufficient. Many other issues related to the situation management should be considered to address completely the SA demons; these issues include all the operations for supporting situation awareness, prediction, reasoning, and control. For this purpose, researchers have proposed a plethora of models, techniques and tools to exploit various research areas and approaches. Despite all these efforts and the valuable solutions provided, the problems of SA demons still remain challenging, and these problems become even wider as the complexity of the systems and the environments increases.

It seems evident that computational approaches and techniques can be effective to support the process of SA assessment if they are strongly founded on solid cognitive models able to exploit and take into account the characteristics and processes of human cognition. Such computational approaches should support the process of SA in all its phases – namely, the perception of the elements in the real environment, the understanding and the identification of the current situation, and finally, the prediction of future evolutions of the situation. Such approaches should facilitate decision-making, and these approaches should directly and explicitly tackle the SA demons. For this reason, these approaches should be defined by addressing the actual motivations that are at the basis of the birth of such demons in real-world applications.

Following this consideration, the overall objective of this thesis is to contribute to the resolution of the SA demons through the study of cognitive models and computational approaches. Specifically, the aim is to identify novel information paradigms for an original support to the SA and to define new theoretical and practical approaches for the resolution of some of the SA demons using computational models and techniques.

To pursue this overall objective, this thesis addresses the following specific research challenges in SA:

i) Users of SA systems may encounter low quality or missing sensor data. The reliability of data strongly influences the SA as it undermines the first level of SA, namely the perception of the elements from the environment. The complexity and the data overload are the main demons hindering this issue.

ii) Users may deal with data overload, attentional tunneling, and misplaced salience demons that affect all the levels of SA. These three SA demons are strictly correlated and they mutually influence each other. Due to these demons, users may not be able to identify the most important elements of the environment and of the system which they should attend to.

iii) Wrong mental models may be used, and this is at the basis of many errors in SA. Having a wrong mental model prevents correct interaction with the system. Users need approaches and techniques that promote the use (and the building) of correct and effective mental models.

iv) Effective situation models need to be defined. The representation of the situations plays a crucial role in supporting SA. Users should be able to see and directly interact with the representation of the situation; they should also be able to see it from different points of view and to reason on the possible evolutions of the situations in the future.

To address these research challenges in achieving the overall goal, this thesis identifies the following research objectives and the activities and tasks needed to fulfill them:

Objective 1. To study and analyze the current state-of-the-art approaches and techinques in terms of solutions for the SA demons:

- **Task 1.1.** Understand the impact that the SA demons has on the SA assessment processes and which are the main factors and elements that influence (positively or negatively) the SA demons.
- **Task 1.2.** Analyze the main research works in Computer Science that contribute to the SA.
- **Task 1.3.** Realize a mapping between the main research focus areas of the analyzed works and a functional view of an SA system: this allows the

identification of the most common approaches used in each phase of the process of SA from a computational perspective.

Objective 2. To investigate the usefulness and effectiveness of Granular Computing (GrC) for SA:

Task 2.1. Examine the GrC paradigm thus to identify the commonalities between GrC and SA.

Task 2.2. Conduct a preliminary feasibility study for the application of some features of GrC in SA.

Task 2.3. Conduct a thorough analysis of the state-of-the-art techniques of GrC, and develop demonstrators and prototypes which implements such techniques to identify which phases of an SA system they can support effectively.

Task 2.4. Define a new theoretical framework for supporting perception, comprehension and projection of situation based on GrC.

Task 2.5. Define an original model of situations based on GrC which can support the users in dealing with the SA demons: attentional tunneling, data overload, complexity, workload and stressors, wrong mental models.

Objective 3. To deal with the problem of low-quality and missing sensor data by defining a new computational approach, which can be useful to address the data overload and complexity demons at the perception level:

Task 3.1. Define an original approach for the management of sensor data with the concept of virtual sensors.

Task 3.2. Define a novel sensor data imputation technique based on association rule mining.

Objective 4. To define an original computational approach to situation management that is able to deal with attentional tunneling, data overload, wrong mental models and misplaced salience demons:

Task 4.1. Define an approach for supporting the trade-off between goal-driven and data-driven information processing as a valid solution for the considered demons.

Task 4.2. Define an approach for suggesting the information on which the user should pay attention to limit the issue of data overload and attentional tunneling.

Task 4.3. Design and develop prototypical systems for implementing the proposed approach and evaluate them in real-world applications.

1.2 Approach

The objectives delineated in the previous section are carried out by conducting this study as described in Figure 1.1.

The first part of the research deals with the **study of the SA models, errors and demons** in order to identify the main factors that hinder the process of SA assessment. This step provides the sufficient theoretical background to study and analyze the approaches, models, techniques and systems of SA. This analysis is conducted with the aim of realizing an **overview of the state-of-the-art** in the area of computer science that addresses the issues of SA. Specifically, the overview is conducted according to the main research focus areas (e.g., ontology, data mining, computational intelligence, and architectures), which the research works mainly refer to. The results of the overview provide greater insight into the main research areas that are used for implementing and supporting the main phases of an SA system. To this purpose, this thesis realizes **a mapping between the main research works and the phases of a functional view of an SA system**.

Although some techniques belonging to the GrC paradigm (e.g., fuzzy logic, rough sets, clustering) are already adopted to support some functionalities of SA systems, such techniques are not used by exploiting entirely the potentiality of the GrC paradigm. This means that the SA systems are not benefitting from the support that it may give for the data analysis and decision-making. Thus, a deep **study and analysis of the GrC paradigm** is conducted; in this step, the different commonalities between the principles and characteristics of GrC and SA are identified. Some preliminary verifications on the applicability and the usefulness of GrC techniques to some features of SA systems are conducted. After this analysis, it is possible to define a **mapping between the main research focus areas of GrC and a functional view of SA system**, thus extending the previous mapping.

After this initial phase of analysis, we focus on the main research challenges regarding SA demons. To address these research challenges, we need to leverage on some consolidated and widely exploited research approaches and techniques, such as **data mining** (to address problem at the perception level) and the **ontologies** (to support the situation management). These need to be used in an original way to define new and effective approaches. In addition, we investigate and exploit a new information processing paradigm – namely GrC – to propose an innovative and original support to the whole process of SA assessment.

With respect to Data Mining approaches, we exploit a rule mining technique to support the SA at the perception level by addressing the problem of low-quality and missing data. In particular, the aim is to define a **quality-aware sensor data management approach**, based on virtual sensors, which implements a sensor data imputation technique based on association rule mining.

Regarding the ontological approaches, the aim is to define a novel **adaptive goal-driven approach for the situation management** (AGSM) which addresses the demons of attentional tunneling, data overload, misplaced salience, and wrong mental models. The approach is implemented in three different prototypical systems in the domains of e-learning, fleet logistics, and port logistics.

Furthermore, the GrC paradigm is used to define a **theoretical framework based on set theory** for supporting all the phases of SA. The rough sets formalism is exploited to define a **model for representing situations**.

Lastly, the research approach foresees the evaluation of the proposed contributions. Studies and analysis are conducted to identify the best **approaches and techniques to evaluate the contributions**. Wherever possible, we use numerical and quantitative evaluation methods. Specifically, the quality-aware sensor data management approach is evaluated by measuring its accuracy on a real dataset. The AGSM approach is evaluated by using the Situation Awareness Global Assessment Technique (SAGAT) with real users [19] and using the three prototypical systems we have developed. We evaluate also the capability of the approach to improve the performances of decision making processes by means of numerical simulations in the domain of port logistics. Lastly, using two different demonstrators, applied to the maritime and aviation domains, we validate the usefulness of the set-theoretical framework for SA and the model of situations based on Rough Sets.

Figure 1.1 depicts the approach used in this thesis to achieve the research objectives and describes the relations between the research activities, objectives and tasks.

In the next section, we briefly summarize the main research contributions of this thesis.

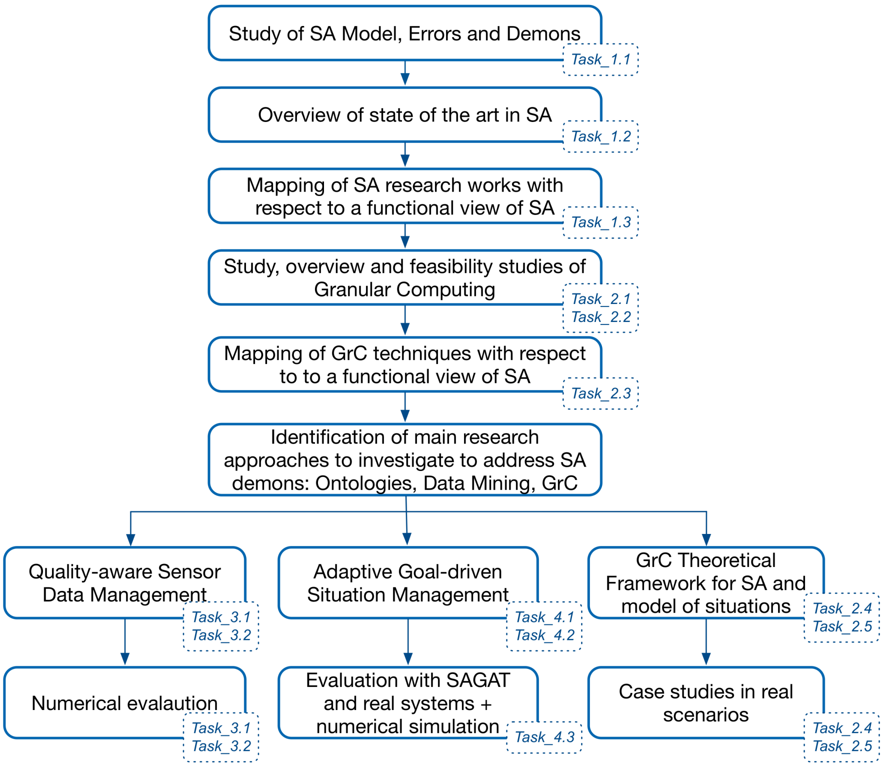


Figure 1.1 Research approach.

1.3 Contributions

This thesis contributes to the resolution of SA demons. This section outlines the main contributions, which are listed as follows:

- An overview of the state-of-the-art in terms of approaches, techniques, models, and systems for SA, classified according to the main thematic research areas to which they refer and a mapping of the research works with respect to the functionalities of an SA system.
- Identification of the main commonalities between the GrC paradigm and the SA paradigm.
- A mapping computational techniques of GrC and the main functionalities of an SA system they support.
- A quality-aware sensor data management approach which addresses the demons of data overload and complexity at the perception level of SA.
- A sensor data imputation technique based on association rule mining.
- The Adaptive Goal-driven Situation Management (AGSM) approach – a novel approach for supporting situation management based on the active goal of the users – which consists of:
  - a semantic model of users goals and situations that sustains the entire approach;
  - an approach for suggesting the most desirable goal to the user;
  - and a reinforcement learning technique that adapts the process of goal selection to the user feedback.
- A set-theoretical framework of GrC to support the whole process of SA assessment.

- A model for representing situations based on rough sets.

- The design and implementation of the following: three prototypical systems (system for supporting self-regulated learning, the green fleet management system, the DSS for supporting port logistics operations); a discrete-event simulation model of port container terminal operations; two demonstrators of the GrC theoretical framework and of the situation model based on rough sets; and a demonstrator of the quality-aware sensor data management approach.

Regarding the evaluation of such contributions, we propose using:

- a numerical evaluation of the data imputation technique of the quality-aware sensor data management applied to a real dataset using the developed demonstrator;

- an extensive evaluation of the AGSM approach in three different domains using the SAGAT methodology and the developed prototypes;

- a numerical simulation of the logistics operations of the port container terminal of Salerno, Italy, to quantify which is the impact of the AGSM approach on the performances of the operations, using the developed discrete-event simulation model;

- a case study regarding the movement of vessels in the harbor of Salerno, Italy, to demonstrate the support of the model of situations based on rough sets for supporting SA using the developed demonstrator; and

- a case study in the aviation domain, regarding the splitting maneuver, to demonstrate the usefulness of the set-theoretical framework in supporting expectations about situations by means of a conformity analysis technique implemented in the developed demonstrator.

1.4 Thesis Outline

Chapter 2 introduces the theoretical background on Situation Awareness, by describing the model adopted as a reference in this thesis and examines the Situation Awareness Demons. Then, the chapter presents the overview of situation awareness approaches, techniques, models and systems, classified according to the research focus area to which they refer. The chapter concludes with a mapping between the analyzed research works and a functional view of a generic SA system. Chapter 3 introduces the Granular Computing paradigm. It describes the main commonalities between the GrC and SA on a philosophical, methodological and technical level. Then, an overview of the computational techniques of GrC with respect to their support to the phases of SA is described. Lastly, a mapping between the main research focus areas of GrC with a functional view of the SA is proposed. Chapter 4 proposes an overview of the main contributions of the thesis, that leverage and are grounded on the results of the analysis reported in the previous two chapters, with respect to the situation awareness errors and demons they address. Moreover, some issues on the evaluation of SA are discussed, thus presenting the evaluation methods adopted in this thesis. Chapter 5 describes the quality-aware sensor data management approach. Details about the sensor data imputation technique based on association rule mining are provided. The chapter is concluded by presenting the evaluation results obtained applying the technique on a real dataset. Chapter 6 describes the Adaptive Goal-driven Situation Management techniques. Specifically, after a description of the overall computational approach, details about the semantic model of goals and situations and about the goal selection mechanism are provided. Then, the chapter describes the reinforcement learning technique that makes the approach adaptive to the users' feedback. The three prototypical systems that implement the AGSM approach are herein described, together with the results of the evaluation obtained using SAGAT. Moreover, the numerical simulation to demonstrate the improvements in the performance of operations and decisions is described. Chapter 7 introduces the set-theoretical framework of GrC for supporting SA. It presents the novel concept of representing situations as granular structures. Then, it describes a conformity analysis technique based on such framework for dealing with users' expectations about situations. The chapter describes also the novel situation model based on Rough Sets and its application to the case study of vessel movements. Lastly, an evaluation of the framework by applying the conformity analysis to a scenario in the aviation domain is provided. Chapter 8 discusses the future work and challenges and concludes the thesis.

Chapter 2

Situation Awareness

> *"Every man takes the limits of his own field of vision for the limits of the world."*
>
> — Arthur Schopenhauer, *Studies in Pessimism: The Essays*

In this chapter, we recall fundamentals theoretical background concepts concerning situation awareness (SA). First of all, we briefly describe the main elements of the Endsley model of SA, a widely adopted conceptual model focused on the relations between situation awareness and dynamic decision making. Leveraging on the works of Endsley and other scholars, we analyze the main common errors affecting the process of SA assessment and consequently degrading the decision making performance. We discuss the main functionalities and capabilities a computational system should possess to effectively support SA. According to this functional view of an SA system, an overview of existing approaches, techniques, methods and models implementing each phase of an SA system is proposed. The analyzed solutions have been classified according to the research focus area they refer to, obtaining a systematic view where approaches and techniques are mostly adopted for a given functionality of an SA system.

2.1 Theoretical Background

Situation Awareness (SA) is a faceted concept encompassing many different elements ranging from cognitive mechanisms and decision making processes to information processing and human factors. Consequently, providing a universal definition of SA fitting for different contexts is not an easy task. Intuitively, SA means to understand what is happening around us in a specific moment in order to be able to perform a correct action or make a coherent decision with respect to our goal. Many formal definitions of SA have been proposed since the late 1980s, when the concept of SA started to gain a growing attention in the military aviation context [20], [21], [22], [19]. In subsequent years, the concept of SA started to be applied in a plethora of different domains and contexts where the human operators have to make timely decisions or operate on a complex system [23], [24].

This thesis focuses on an operational view of the situation awareness, which means that people must have SA for a specified reason, i.e., to complete a task. Specifically, we aim to understand the role of SA in the interaction of humans with dynamic and complex systems. In such systems, decision making and performances strongly depend on the operator's situation awareness.

Our perspective is that SA incorporates an operator's understanding of the situation as a whole, forming a basis for decision making. From this perspective, the definition that best fits with our goals is the one provided by Endsley in 1995 [1]:

"*Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future*".

In this definition we can identify three levels which concur to the formation of the SA: perception, comprehension and projection, ad described in what follows:

- **Level 1 SA: Perception**: the first level of SA (level 1 SA) is the perception of the status of the elements in the environment. The elements to which pay attention depend on the task to be performed. A driver of a car should perceive the other cars on the street, the characteristics of the street, the status of the car, eventual pedestrians. All the senses can be used to perceive the environment from different and heterogeneous sources. Although it can seem easy to perceive the elements related to a specific task, in many domains it can be quite challenging just to detect all the needed data. Moreover, quite often the amount of data rapidly outpaces the capability of the operator to correctly perceive all of them. Jones and Endsley found that most of the SA errors (more than three-fourths) occur at level 1 [15].

- **Level 2 SA: Comprehension** The second level to achieve good SA (level 2 SA) is understanding what the data perceived at level 1 mean in relation to goals. The elements of level 1 are synthesized and aggregated and then a goal-related meaning is associated with each piece of data. Understanding the meaning of the perceived data requires a good knowledge and a good mental model in order to put together and interpret different pieces of information.

- **Level 3 SA: Projection**: Level 3 SA means to predict what the perceived elements will do in the future with respect to the goal. Level 3 SA depends on the correct understanding of the situation (level 2 SA) and on the knowledge about the dynamics of the system and of the environment. Usually, such operation is quite demanding, as it requires a good understanding of the domain, of the situation and a great ability in the projection of the status of many elements in the future. Experience plays a major role in this level because it gives the ability to anticipate future situations and to be proactive with respect to them.

The three levels of SA should be interpreted in the context of a dynamic decision making process. Indeed, Endsley proposed this definition in the set of a wider model of dynamic decision making (depicted in Figure 2.1) in which the situation awareness is a previous, separated stage from the decision making. According to this

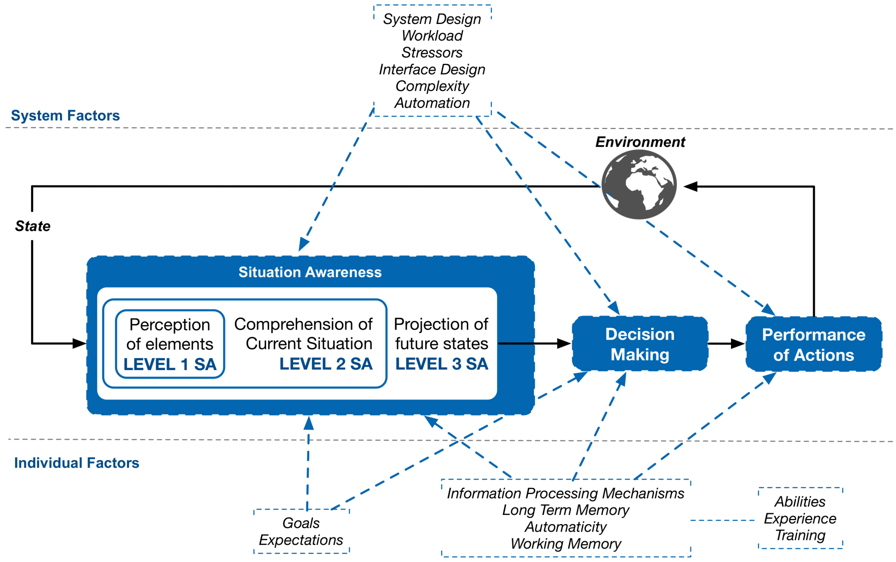


Figure 2.1 Endsley's Model of Situation Awareness in dynamic decision making (our re-elaboration from [1])

representation, the SA is the operator's internal model of the state of the environment: the operator makes a decision according to what is going on in the environment (i.e., the identified situation) and what he/she thinks will happen in the near future and thus he/she performs the necessary actions. Such actions will produce an effect on the environment, thus modifying the situation, entailing a new cycle of situation awareness and decision making. An important distinction between the situation awareness intended as the internal model of the world, i.e., a state of knowledge about what is happening in a given moment, and the process of gaining and maintaining the situation awareness by elaborating and understanding new information. We refer to the process of gaining situation awareness as *situation assessment*.

2.1.1 Cognitive Mechanisms influencing Situation Awareness

The SA is the product of different cognitive processes and it is based on cognitive mechanisms that we briefly describe in what follows. Further details can be found in [1, 25, 15, 26].

The *Working Memory* (WM) serves to store information in the short period. WM may contain only a limited amount of information (7 ± 2 pieces of information). Maintaining information in the WM demands for a huge cognitive effort, otherwise such information will decay. The limited amount of WM may rapidly cause a serious loss of SA when many information needs to be perceived.

The *Long Term Memory* (LTM), instead, consists of different structures that store information for long time, overcoming the limitations of the working memory. The *mental models* and *schemata* are two structures of the LTM which play a major role in SA. Mental models can be thought as structures which model the behavior of specific system (physical or abstract system), its purpose, its possible (current and future) states, its function, how it works. Mental models are crucial to SA because they support people in the identification of the information to pay attention to, and they help in the creation of *expectations* of what can happen next. With poor or wrong mental models, a person has very few chances to understand what is happening and what will happen next. Another construct of the LTM is the *schema*. A schema is a prototypical system state which represents the system in a certain situation. By means of pattern matching processes, specific information perceived from the environment can be matched with the available schemata in a short time, to identify which is the best schema that match the current situation. This mechanism represents a further improvement in term of processing efficiency with respect to the mental model, allowing a person to rapidly classify and understand a well-known situation that is reoccurring. The last construct of LTM is the *script*, a sequence of actions to perform in a given situation: having identified the best schema that matches with the perceived cues, it is possible to directly perform the actions of the related script, thus speeding up the process of responding to what is happening.

A fundamental cognitive mechanism is represented by *goals* that drive all the process of SA assessment and decision making. In the so-called *top-down information processing* or goal-driven information processing, the goal identifies the information of the environment to which pay attention. Indeed, the *active* goal identifies the specific mental model the operator will use. Thus, it is important that the people always focus on the right goal: when focusing on the wrong goal, important information may not be correctly perceived and interpreted.

Individuals have also some **expectations** of what could be the state of elements (even without directly perceiving information about such elements) and what will happen next. When having expectations, people do not need to acquire and process other information from the environment. Unfortunately, false expectations can lead people to misinterpretations of data and to miss some information.

Lastly, **automaticity** allows to react to what is perceived without using mental models or other cognitive processes, thus freeing up many mental resources for other tasks. Unfortunately, in some situations, automaticity can be dangerous for the SA as the information that are outside the loop of the automaticity may not be perceived.

2.1.2 System Factors influencing Situation Awareness

The cognitive mechanisms described in previous section represent the individual's factors that intervene in the process of SA assessment. Besides such individual factors, the following task and system characteristics influence the SA (with both positive or negative consequences) [1]. First of all, the **system and interface design** play a major role in influencing the SA. In the various phases of information processing, from the environment to the operator, something can go wrong, affecting the SA. The system may not gather all the needed information, due to issues with the sensors. Then, not all the information acquired can be displayed on the interface. Lastly, even if the information is shown, it can be only partially transmitted to the operator due to the limitations of the cognitive mechanisms. Systems and interfaces for SA should be designed trying to reduce the mental workload. Jones and Endsley in [15] provide some principles for designing systems and interfaces.

Another system factor influencing SA is the **stress**. Some stress factors are related with the system, like physical stressors (noise, vibration, heat, etc.) or social stressors like fatigue and uncertainty. A certain amount of stress can be beneficial as it usually improves the degree of attention. Higher amount of stress has negative consequences on the SA.

Related with the concept of stress, also the **workload** can decrement the SA, as a high workload can outpace the cognitive capabilities of humans. System **complexity** affects negatively the SA because it usually increases the mental workload. The complexity of the system can be soften both with a proper design and by training the operator in order to create a proper mental model.

The level of **automation** of a system can help in reducing the workload, but when too automation is available, the out-of-the-loop syndrome may happen, causing a loss in the SA.

2.1.3 Situation Awareness Demons

The coexistence and interaction of cognitive mechanisms and system factors contribute to hinder the process of SA assessment, especially in those domains where systems are complex and a great deal of information have to be processed in a short time. Such individual and system factors give birth to a set of issues with the SA assessment, namely the "SA demons "[15], which should be taken in serious consideration when designing an SA system. We briefly report the definition of each SA demon by leveraging on the work of Endsley [15].

- **Attentional tunneling**: To have a good SA, a human operator should share the attention between different tasks and information sources. When the operator narrows the attention only on a subset of all the available information, the phenomenon called *attentional tunneling* happen: the operator remains stuck on one single task, losing the view on the global picture, with a critical loss of SA.

- **Data overload**: In many domains, the amount of data that needs to be processed and the rate at which such data changes, rapidly overwhelm the cognitive capabilities of the operator. A sapient design of the system, together with the adoption of suitable data analysis techniques, represent powerful means to reduce the issue.

- **Complexity**: When the complexity of a system (or of a task) is too high, with many features that are not straightforward and many complex rules that govern the functioning, the operator is not capable of constructing a correct mental model. Training and experience can be useful to solve this issue, as well as trying to design interfaces able to simplify the interactions.

- **Memory trap**: The limited space available in the working memory and the huge cognitive effort to maintain information in it, undermine the SA in those situations in which the operators need to remember many information. The use of correct mental model, as well as the experience, may help in freeing up the working memory.

- **Workload and stressors**: Excessive workload, stress, fatigue, anxiety, undermine the SA seriously. The sources of these stressors are multiple, from environmental to task to individual and personal factors, and they reduce the efficiency of the working memory.

- **Wrong mental models**: When a poor or wrong mental model is adopted, the operator fails in understanding the meaning of the perceived information and in projecting the situation in the future, thus making wrong decisions. This often happens when users are in front of a new system but they try to adopt a mental model that was correct for the old one.

- **Misplaced salience**: The capability of the elements of the interface to properly catch the attention of the users on the important information (e.g., alarms, unexpected events, dangers) is defined as *salience*. Salience of information depends especially from the physical characteristics of the elements (e.g., red color, blinking lights, alarms). The salient properties help to increase the level of SA if properly exploited, but in cases where too much elements require the operators' attention, it causes the opposite effect.

- **Out-of-the-loop**: Automation can help people in maintaining high level of SA as it reduces the workload. Unfortunately, when the automation fails, the out-of-the-loop syndrome may happen: people usually rely on such automation to perform the task, and so they are out of the loop of control of the system. In such circumstances, they are not able to detect a problem and to intervene in a timely manner.

2.1.4 Situation Awareness Errors

The SA demons described in the previous section are among the main factors that affect the SA assessment process. Their presence causes an incomplete (i.e., knowledge of only some of the elements of the environment) or inaccurate (i.e., erroneous knowledge of the elements of the environment) SA. These demons lead to different kind of errors at all the three levels of SA. In [14] a taxonomy of the most common errors in SA is proposed. In what follows, we briefly describe such errors, by evidencing also their relations with the SA demons.

At level 1 SA, the errors are related to the lack of perception of important information. The main SA demons that contribute to such errors are attentional tunneling, data overload, memory trap, workload and stressors. The most common errors are:

- Data not available due to a bad system design or to a failure in the communication process.

- Data difficult to perceive due to system or environmental issues.

- Failure to observe data due to: i) omission of the operator that fails in properly setting the system ii) attentional tunneling; iii) high workload.

- Misperception of data, mainly due to the misuse of expectations that leads the operator in misperceiving a given element from the environment. In other cases, the operator may simply misread the data (e.g., reading errors, spatial disorientation).

- Memory failure due to the lack of memory and data overload

Errors at level 2 SA relates with the failure to comprehend the situation even when the information have been correctly perceived at level 1. Among the others, the main SA demons that contribute to these errors are the wrong mental model, the attentional tunneling and the memory trap.

- Poor mental model leads to a wrong understanding of the meaning of perceived information

- Wrong mental model, i.e., the operator adopts the mental model of a system to interpret the information of a different system (e.g., a driver that drives a new car that has a different disposition of switches and indicators). This cause a misinterpretation of the current situation.

- Over-reliance on default values in model, in the absence of real-time data, can cause a serious loss of SA when the default values do not correspond to the actual ones.

- Memory failure hinders the process of integration of all the perceived information.

At level 3 SA, errors happen when the operator is not able of predicting the future state of what has been understood at level 2 SA. This happens mainly due to the presence of poor mental model of the system. Projecting situations is a very demanding task, which usually requires experience with the system.

Lastly, other two factors impact negatively on all the three levels of SA. In order to have a global picture of what is happening in the environment, it is needed to manage and *maintain multiple goals* in memory and perform multiple tasks. Some operators may be poor at maintaining multiple goals especially for attentional tunneling, misplaced salience and memory trap. Moreover, the use of *habitual schema* leads operators to be less receptive to important environmental cues, mainly due to the attentional tunneling demon.

2.2 Situation Awareness Systems: an overview

In the previous section we have delineated the main factors that hinder the process of SA assessment in dynamic decision making context. Many intelligent systems, based on different approaches and techniques, have been proposed so far to support human operators in gaining and maintaining SA in complex and dynamic environments [27]. In this section, we propose an original overview (without any claim of exhaustiveness) of the main approaches, methods, techniques and models proposed for the realization of SA systems. The objective of such overview is to identify and classify the main research focus areas that contribute to the defi-

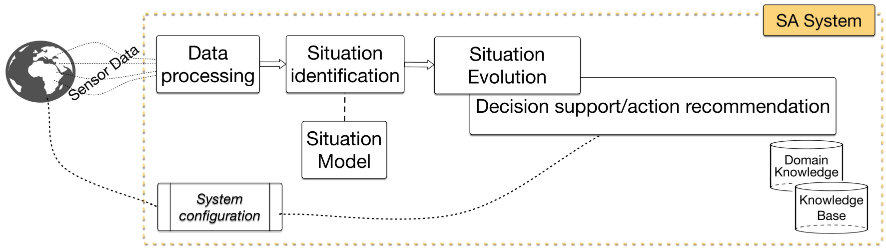


Figure 2.2 A conceptual view of a generic SA system.

nition of SA systems, offering an overview of the contributions at the state of the art.

First we propose a functional view of an SA system, in terms of its core components and functionalities, which is based on the Endsley model, as depicted in Figure 2.2. Such figure is our re-elaboration of the systemic viewpoint of automated SA system proposed in [27]. The aim of an SA system is to support operators in gaining and maintaining SA [27]. With reference to Figure 2.2, we can identify the following functionalities of the system.

- Processing input data: the system gathers the data from the environment and represents it according to a specific data model. In this phase, all the problems related to data pre-processing, data cleaning, data standardization, data reliability should be addressed. A specific issue to address in this phase is the incompleteness and inconsistency of data, which may lead to erroneous comprehension of the situation in the next phase.

- Situation identification: the data gathered by the system is processed according to a situation identification technique in order to assess the current situation. Usually, a formal and explicit model of situation is adopted.

- Situation model: a model of situation is needed in order to provide a concrete support to SA, as well as to support the other capabilities of the system. The support to decision making, the interface updating, the adaptation of the system, the notification of alarms and events, depend on the identified situation.

- Decision making and action support: the role of situation awareness is essentially to support decision making and the consequent performance of actions. In some applications, it is also possible to link the identified situations with suitable actions to perform (even automatically).

Beyond the process of gaining SA, the systems should implement processes for maintaining SA. This essentially requires the following features:

- Tracking evolving situations: after having identified the situations, it is essential to track its evolution over the time, supporting the operators in the monitoring of such evolution.

- Supporting projection: SA system should provide the means for supporting the operators in anticipating and forecasting the possible evolution (level 3 SA) of the situation.

- Adaptation capability: the SA system should be able to adapt itself (i.e., to adapt the different functionalities) according to the changing situations, in order to effectively respond to the users' requests.

Using this functional view of an SA system, in what follows we propose an overview of techniques and approaches for the different phases and features highlighted above, and specifically:

- Data processing techniques, related with those techniques that address the problem of data preprocessing and low-level information fusion.

- Situation models and situation identification techniques: we describe together models of situation and techniques for situation identification because the latter strongly depend on the way the situation is represented.

- Techniques and approaches for supporting projections, i.e., those techniques that explicitly address the problem of situation evolution over time (including the tracking of such evolution), in order to help operators in forecasting the future state of the environment.

- Techniques and approaches for supporting decision making: approaches and techniques aiming at supporting decision making.

Lastly, we analyze complete frameworks for SA which usually integrate different techniques and approaches. We distinguish between general purpose frameworks (which can be used for different applications) and domain-specific systems (which have been designed for a specific application).

2.2.1 Data Processing Techniques

Defining techniques and approaches for data processing is one of the first step in the realization of an SA system. As aforementioned, most of the problem of SA are at Level 1. This is due to the huge amount of data to process, but also to the characteristics of such data: heterogeneity, uncertainty, unreliability, noise. Besides the traditional approaches of data preprocessing to avoid the so-called "garbage-in, garbage-out effect " (e.g., data cleansing, outlier detection, data representation, data normalization, feature extraction), some specific techniques directly address the issues related with SA, like data reduction (to avoid overload), improving data reliability, events detection, and so on. Many works in this area relate with the data fusion theory. Zhang et al. [28] propose the adoption of the Dempster-Shafer Theory of Evidence (DST) [29] to fuse data representing security alert conditions gathered by heterogeneous sensors deployed in a computer network. Wu et al. [30] define a generalizable sensor fusion architecture based on the DST to sustain human perception with capabilities to deal with uncertainty in data and inference mechanisms. Due to the proliferation of smartphones with heterogeneous sensing capabilities, some works consider the smartphone as a valuable source of information and propose approaches to fuse information gathered by such devices to improve SA [31, 32]. In [33], an approach for the identification of possible car accidents based on smartphone data is proposed, in which contextual information are used to avoid false positives. Many approaches use smartphone data to recognize daily-life activities, as in [34, 35].

An important capability to sustain SA at level 1 is the recognition of events or the identification of higher-level features and objects by abstracting and classifying low-level data. Many works adopt artificial neural networks (ANN), as in [36] where an approach based on ANN for the identification of environmental hazards using data coming from different sensors is proposed. Coronato et al. in [37] define an approach for the preprocessing of sampled signals gathered by sensors in order to get a set of statistical features. Such features are exploited by means of an ANN that classifies the temporal frames and generates events for abnormal statuses. The results of the classifier are further improved by means of intelligent agents implemented in PROLOG.

In [38], the authors exploit Markov Logic Networks (MLN) to encode uncertain knowledge and fuse data coming from multiple (and possibly heterogeneous) sources, and perform reasoning on incomplete data to support situation awareness in the maritime domain. MLN combines the power of first-order logic and the probabilistic uncertainty management of Markov networks. With this approach, it is possible to combine different types of knowledge with associated uncertainty for the identification of complex events as a logical combination of simpler evidences. [39] proposes a data abstraction technique based on Symbolic Aggregate Approximation (SAX) to create patterns from sensor data. The created patterns are then linked to semantic descriptions that define thematic, spatial and temporal features.

Another area of great interest for the SA is the use of contents generated by social media as a valuable source of information to increase the SA of decision makers. Salfinger et. al in [40] propose a semantic-based crowd-sensing system for crisis management in which they extract knowledge from Twitter data about a natural disaster. In [41] the authors analyze the spatial and temporal characteristics of the Twitter feed activity during the earthquake on the East Coast of the US in 2011. Such feeds are considered as a hybrid form of a sensor system that allows for the identification of the impact area of the event. Basu et al. in [42], instead, highlight the unstructured nature of the content on social media and thus they criticize their usefulness in post disaster management. For this reason, they propose a framework, based on interactive crowdsourcing, in which an automated system is able to sustain a dialogue with the members of a crowd to collect useful information that helps to build a structured repository of situational information. Vieweg et al. in [43] conduct a study about two emergency events in North America in which they analyze the microblog posts on Twitter in order to identify information that may contribute to Situation Awareness. Saleem et al. in [44] propose an adaptive filter to obtain information related to disaster management from Twitter, trying to reduce noise in the data. After having filtered the information, they classify the tweets as disaster related/unrelated by means of a Naive Bayes algorithm. The authors in [45] address the problem of gathering high volumes of accurate data in disaster management situations. This work proposes an ad-hoc wireless architecture that support disaster response with distributed collaborative sensing.

2.2.2 Situation Models and Situation Identification Techniques

In Section 2.1 we provided a definition of what is a situation from a human perspective. For the design of a system supporting SA, it is useful to define the situation also from a computational viewpoint. In order to provide the users with meaningful information related to their goals, the pieces of data gathered by the sensors should be interpreted into a higher, domain-relevant concept. This higher-level concept is called a *situation*, which is an abstract state of affairs interesting to applications [46]. Such situations provide a simple, human understandable representation of sensor data to applications and thus to the users, whilst shielding them from the complexities of the low-level data. More formally, a situation is defined by Ye et al. as the *external semantic interpretation* of data [46]. Many different techniques for situation identification (i.e., the process for deriving a situation by interpreting and fusing pieces of information) have been proposed, together with the corresponding situation models. According to Ye et al. [46], such techniques can be classified in two categories: i) *specification-based approaches* relying on the representation of expert knowledge about the situations and on reasoning techniques to infer proper situations from data; ii) *learning-based approaches* which exploit machine learning techniques.

2.2.2.1 Specification-based Approaches

Traditionally, specification-based approaches are the most widely adopted, especially when the number of sensors and the number of relationships among the interesting events is low.

Formal Logic models of situation and first-order logic approaches have been widely used to represent situations. The goal of formal logic approaches is to provide a theoretical foundation for an SA system. For instance, situation calculus is a logic language that formally represents a situation as a snapshot of the world at a given instant [47, 48]. Levesque et al. in [49] propose a logic programming language based on the situation calculus called GOLOG. Situation calculus has been applied also in other domains, like in business information systems [50] or in ambient assisted living as in [51]. Situation theory, formerly introduced by Barwise and Perry in 1980s, is a mathematically based theory of natural language semantics [52], further developed by Devlin in [53]. The situation theory provides a set of mathematically-based tools to analyze the way context facilitates and influences the rise and flow of information [53]. Mechkour in [54] provides an overview of the applications of situation theory to model context information and its use in computer science. Dapoigny and

Barlatier proposed the Situation-based Dependent Type Theory (S-DTT), which is an extension of situation theory in order to overcome its limitations concerning the expressiveness and reasoning capabilities [55]. S-DTT is based on the Extended Calculus of Constructions (ECC), a theory used in software validation as well as in mathematical formalization. As pointed out by Ye et al. in [46], many approaches rely on predicate logic for defining situations, as in [56, 57]. Such works use logical formulae in order to define situations by abstracting sensor data. Other works, like [58, 59], propose a declarative approach to represent and identify situations, usually exploiting logical programming language like Prolog.

Another category of expert-based approaches are those based on **evidence theory**. Dempster-Shafer theory (DST) of evidence is one of the most used mathematical theory of evidence [60]. The advantages of using DST rely on the explicitly representation of the relations between sensors and situations, the representation of multiple types of uncertainty and a human-understandable model of situations. However, usually huge effort for modeling situations with DST are required, involving also domain experts. Moreover, only rarely a same model can be directly applied to different domains and applications. For these reasons, DST is not used in complex domains where many situations exist. McKeever et al. [61] propose an approach for situation inference based on DST that takes into account the uncertain nature of sensor data. [62] proposes an extension of DST for supporting situation reasoning in ubiquitous computing environments. [63] focuses on the identification of activities, which can be considered as a special case of situation, using evidence theory that incorporates temporal reasoning.

Another formal approach for situation identification and reasoning is the Context-Space Theory (CST), initially proposed by Padovitz et al. [64] and further developed by Boystov and Zaslavsky [65, 66]. CST is based on a multidimensional space that allows a clear insight on the identified situation.

Other approaches exploits **fuzzy logic** to deal with the uncertainty and vagueness of sensor data. In the most simple approaches, fuzzy logic is used to map sensor data to linguistic variable that make sense to users [67, 68]. Specifically, Anagnostopouolos et al. [68] use fuzzy inference in order to evaluate a situation that is most similar to the current unknown situation by evaluating the similarity between the specification of a situation and the sensor data. Zaho et al. propose a model of situation based on fuzzy sets for enhancing security systems [69]. A fuzzy inference approach is used to abstract the information coming from sensor data in order to verify a potential dangerous situation. Furno et al. [70] proposes a fuzzy ontology-based approach to model situations and to support situation reasoning, thus combining the formalism of ontology with the capability of fuzzy logic to deal with vagueness and uncertainty. Jones et al. [71, 72] propose a situation identification technique based on Fuzzy Cognitive Map (FCM). Also Chandana et al. propose the use of FCM for supporting situation assessment in coastal surveillance applications [73].

Ontological models are among the most used approaches for modeling situations, thanks to their capability of formally and explicitly representing the situations and the domain knowledge, making such models understandable, sharable, reusable by both man and machines. Pai et al. [74] propose an overview of different approaches based on ontological models. Kokar et al. [75, 76] proposed an ontological model of situation based on the situation theory of Barwise and Devlin, namely the Situation Theory Ontology (STO). Dominguez et al. [77] applied the STO for the realization of a computational framework for identity-based security. On the same topic, Kayes et al. [78] propose an ontological model of situation in terms of access purpose. [79] proposes a general-purpose ontological model of situations and a process for situation identification and projection based on ontological inference, capable to adapt the recognition of situations upon the users personal characteristics, goals, and environment. Meditskos and Kompatsiaris in [80] present the ontology-driven framework "iKnow", in which a model of situation based on OWL ontologies model the dependencies between low-level and high-level activities. Matheus et al. [81] propose an ontology for situation awareness, namely SAW Core ontology, which is based on the concept of relation between situation objects and events that may happen in the environment. [82] proposes an ontological model for hierarchical distributed representation of knowledge in autonomous underwater systems. In [83] the authors propose a ontological framework for situation awareness called “BeAware!”. BeAware!’s ontology introduces the concept of spatio-temporal primitive relations between observed real-world objects thereby improving the reusability of the framework. Due to their flexibility, ontologies have been integrated with other approaches to improve the capability of situation identification. Beside the already mentioned work of Furno et al. [70] which integrates ontology and fuzzy logic, Kokar and Endsley propose a cognitive model of situations based on STO and Fuzzy Cognitive Map. Riboni et al. [84] integrate ontological reasoning with statistical inferencing to improve the process of activity recognition. Metzke et al. [85] integrate ontological knowledge and event processing to support situation identification in the domain of logistics.

Recently, some works adopt the **Complex Event Processing** (CEP) paradigm to support situation identification. CEP is an event processing approach that combines data from multiple sources to infer events and patterns. Stojanovic and Artikis [86] give an overview of the existing approaches applying CEP for real-time situation awareness. Vlahakis et al. in [87] propose a situation model for supply chain management based on CEP in which situations are represented as correlations between simple events, complex events and supply chain objects. Lu et al. in [88] propose an approach for abnormal situations identification based on CEP.

2.2.2.2 Learning-based Approaches

The limitation of specification-based approaches relies on the need of modeling a priori knowledge: in very complex environment with many situations, they are impractical to use. According to the classification of Ye et al. [46], learning-based approaches allow to overcome such issue by leveraging on machine learning techniques that automatically identify complex associations between sensor data and situations, although often such approaches are less precise than the expert-based ones.

Many approaches are based on **Bayesian** classification techniques. Indeed, Naïve Bayes have been used for situation identification with good results, especially in activity recognition [89, 90, 91]. Also Bayesian belief networks have been applied in order to exploit dependencies between low-level attributes and situations [92, 93]. For instance, Wiggers et al. in [94] define an approach for the classification and identification of situations regarding air targets approaching combat vessels by means of Bayesian networks to deal with the uncertainty and incompleteness of sensor data. Morales and Moral in [95] use dynamic Bayesian networks to improve situation awareness of flight crews through the discovery of relationships between real-time flight variables in order to support decisions.

The other family of approaches used for situation identification are based on **Hidden Markov Models** (HMMs) due to their capability of modeling sequence of events. Damarla in [96] defines an approach for situation identification that adopts HMM to exploit data coming from multiple sensors, wherein an HMM represents a sequence of events that leads to a situation. Andersson and Pettersson in [97] defines an algorithm based on HMM for the recognition of aerial-mission via the fusion of information on object tracks. Lison et al. in [98] propose an hybrid model based on Markov logic, i.e. a combination of first-order logic and probabilistic belief network which allows to combine the expressive power of relational structure with the uncertainty of low-level data. Such model is used in a cognitive architecture for human-robot interaction.

Decision trees have also been used as a classification model for identifying situations. In [99] an approach to improve cyber situation awareness by identifying dangerous situations about cyber incidents (e.g., network attacks) is proposed. Lee and Lin propose an hybrid approach integrating decision trees and HMMs to provide a method for identifying situations in smart home environment with low cost sensors [100].

Due to their good performance in pattern recognition, **Neural Networks** have been applied for situation identification [101, 102, 103]. Li et al. in [104] adopt Graph Neural Networks to identify situations involving people in images. Ilin and Perlovsky propose a cognitively inspired mathematical learning framework called Neural Modeling Fields able to identify situations composed of objects while overcoming the combinatorial complexity of associating low-level data with situations [105]. The authors in [101, 106] adopt a fuzzy neural network classifier to detect anomalous patterns in vessels movements and identify dangerous situations by exploiting real-time tracking information. Wang et al. propose an approach based on heterogeneous multi-sensor data fusion using a multi-layer feed-forward neural network to support Network Security Situation Awareness [107].

Another classification approach adopted in the context of situation identification is the **Support Vector Machine** (SVM). The advantage of using SVM relies in their ability of overfitting protection even with large feature spaces. SVM has been used in the context of smart home and pervasive computing for identifying situations related to human activities [108, 109]. In the context of Network Security Situation Awareness, multi-class SVM has been used to identify dangerous situations by processing data about network activities [110, 111]. Lu et al. [112] propose a situation recognition approach based on SVM for supporting decision making with early warnings about given situations.

Some works propose the exploitation of **knowledge mining** and **rule mining** approaches to support situation identification [113, 114]. In [115, 116], techniques of web mining to identify human activity from web data has been proposed. Pournori and Akhgar in [117] investigate the adoption of data mining techniques to improve Cyber Situation Awareness. In [118], the authors provide a survey of clustering techniques exploited for situation identification. Chen et al. [119] propose an approach for situation identification in the maritime domain based on knowledge discovery and genetic algorithm. Dahal et al in [120] define an algorithm for stream mining analysis in order to identify situations and support decision in the domain of smart grids. Many approaches adopt rule mining techniques (especially association rule mining) to identify patterns within sensor data and between sensor data and situations [121, 122, 123, 124].

Recently, some approaches of **Deep Learning** have been applied to the situation identification problem, especially for the ones based on image analysis, although the research in this sense is still at the beginning. Carrio et al. [125] conduct a brief survey of approaches of deep learning applied to situation awareness in UAVs systems. Noever and Regian apply the deep learning paradigm for supporting SA in cyber threats [126]. Tang and Crandall [127] apply deep learning to improve maritime situation awareness.

2.2.3 Supporting Situation Projections

Being able to anticipate the evolution of current situations is at the same time the most important and most difficult task in SA in order to make adequate decisions and perform suitable actions. Most of the situation identification approaches described in previous section provides also some kind of support to the projection phase. In this section, we give a brief overview of those techniques that are specifically designed to support projections. Stojanovic and Artikis provide an overview of approaches based on Complex Event Processing supporting prediction of future situations, with particular attention to the application areas of activity recognition and social media observation [86]. Bomberger et al. define an approach for predicting anomalous behavior of vessels by means of fuzzy neural networks [106]. Multi Entity Bayesian Networks, which is a hybrid approach combining Bayesian Network with first order logic, is applied for predictive situation awareness [128]. Predictive situation awareness is a research area which emphasizes the ability to make predictions about aspect of a temporally evolving situation, which in terms of Endsley's model means to focus on SA level 3 [129, 130]. In [38] the authors propose an approach for reasoning on incomplete data by means of Markov Logic Networks in order to support situation projection in maritime situation awareness for predicting events and anomalies. Fuzzy logic has been investigated to support situation projection. For instance, Zhao et al. in [69] integrate fuzzy reasoning and time series in order to predict future situations about network security and potential attacks and security alerts. Jones, Connors and Endsley [71] provides a data fusion model based on Fuzzy Cognitive Map (FCM) in order to effectively support human cognition, also in the projection phase. The approach is based on the formal representation of user's goals identified by means of the Goal-Directed Task Analysis (GDTA) [15]. Case-based reasoning and domain knowledge has been integrated in [131] in order to support situation prediction while handling uncertainty in data and in the modeled knowledge. Lastly, Naderpour and Lu [36] propose a safety supervisor system based on neural network and expert system to support situation prediction.

2.2.4 Supporting Decision Making

The Endsley's model considers the situation assessment as a previous stage before the decision making processes, highlighting the role that SA owns in making decisions. Thus, many SA systems provide different approaches and strategies to support decision making. In this section, we provide a brief overview of those works that explicitly focus on the decision making by providing an effective support to the users in the context of SA; the approaches and techniques for decision support that are integrated in the context of a wider system are described in next section. Pavkovic et al. [132] provide an overview of decision support approaches based on situation awareness for the response phase of emergency management. Feng et al. propose an approach for supporting decision making by means of a rule-based inference engine able to classify events, recommend actions and support proactive decision making, applied in the context of command and control application

[133]. In the context of smart grids, Dahal et al. define a model able to support decisions in the control room of the grid to improve its reliability by means of stream mining algorithm [120]. In [134], bayesian networks are exploited to support decision making by increasing the level of SA in the domain of safety critical environments. Basu et al. in [42] define a technique of interactive crowdsourcing for supporting decision making in post-disaster management through the interaction with data gathered by social media.

2.2.5 General-purpose Framework for Situation Awareness

In this section, we present an overview of general-purpose frameworks for the realization of SA systems that are not specific for a given domain or application. Salfinger et al. [27] provide a brief overview of SA systems and frameworks , with a specific focus on their support to the evolution of situations. Jacobson et al. in [135] provide a review of major aspects of situation modeling and management, and propose a multi-agent framework for situation management based on event correlation. Baumgartner et al. [136, 83] describe a comprehensive situation awareness framework for control centers based on ontologies. The framework integrates well known ontologies with spatio-temporal reasoning to support the three phases of SA. Furno et al. [70] define an agent-based framework for the implementation of SA systems leveraging on an ontological model of situations and on fuzzy reasoning. In [137], a framework for high-level information fusion based on probabilistic finite state automata and a data fusion architecture for situation awareness is proposed. Brannon et al. in [138] define a system based on POMPDP and ANTMAP [139] for high level information fusion and situation awareness support. The system is able to process different type of inputs and support situation assessment with confidence level and evidence in support and against of situations as output. Cimino et al. [123] define a framework to manage situation awareness by integrating different approaches:

Semantic Web to handle situation inference, fuzzy logic to deal with uncertain data and genetic algorithm to deal with adaptation to user's behavior. The framework is implemented with an agent-oriented architecture to provide a structural interoperability in an open and dynamic environment. Pearson et al. [140] define a framework for SA based on dynamic situation modeling. The Situation Awareness Assistant (SAWA), described in [141, 142] is a tool suite for defining SA applications as rule-based expert systems. It allows to formally model domain knowledge by means of the SAW Core ontology [81]. The system implemented with SAWA can be only useful to detect a priori specified situations, with no capability of generalizing to new, unexpected situations. To overcome such issue, other approaches suggests the integration of knowledge discovery tools and data mining approaches, as in the work of Salerno et al. [143]. In this work, an SA framework helps analysts in the discovery of new domain models and patterns.

2.2.6 Domain-specific Situation Awareness Systems

It can be quite difficult to apply a general-purpose SA system to complex and dynamic environments in which applications have to deal with high number of situations and process huge amount of data. In such cases, domain-specific SA systems can be more effective and useful. Ontologies are widely used for domain-specific SA systems, as in [144] where a system for service recommendation is built upon a semantic model. Also [145] describes a system based on domain ontologies semantic web services for supporting situation awareness in humanitarian emergencies. Pai et al. in [74] describe a high-level information fusion framework based on multi-layer ontology to support military applications. In [78], an ontology-based framework for access control of software services based on situation awareness paradigm is proposed. Another ontology-driven framework is described in [80], whose aim is the recognition of daily living activities. In this case, ontologies are not used as an higher level model of the world, but for representing the dependencies among low-level and high-level objects. The framework has been applied in the healthcare domain. Miguelanez et al. [82] propose a semantic knowledge-based framework for processing distributed knowledge in autonomous underwater system supporting the SA in human-robot interaction.

A domain that traditionally have attracted lot of attention from the SA practitioners is the maritime situation awareness, as demonstrated by many systems proposed in this domain, as for instance in [146, 147, 106, 73, 148]. The book "Situation Awareness with Systems of Systems "[149] describes a complete system for supporting maritime situation awareness (focusing on safety and security issues) by means of the integration of different approaches like multi-objective visualization methods, stochastic outlier selection, rule-based anomaly detection, ontology-based event model. Also Van den Broek et al. in [150] have proposed an ontology-based framework for maritime situation awareness with a technique for multi-sensor data fusion. Chen et al. in [119] propose a system, namely GeMASS, for maritime situation awareness based on knowledge discovery and genetic algorithm.

The management of safety critical environment is another domain in which many SA systems have been proposed. For instance, in [134], a system based on Bayesian networks for supporting SA in chemical plants is proposed. A decision support system which integrates case-based reasoning and context-aware to support situation awareness for the prediction of hydrate formation in gas pipelines is proposed in [131]. By exploiting the Complex Event Processing approach for situation identification, Vlahakis et al. defines a framework for situation awareness in the context of supply chain management [87]. Gariel et al. propose a system based on clustering techniques for the detection of non-standard aircraft landings in airspace monitoring [151]. An advanced driving assistance system based on SA paradigm is proposed in [152] which classifies driver behavior in real time to send alerts in case of dangerous situations.

The management of crisis (e.g., natural disasters, terroristic attacks, plants incidents) can benefit by the application of SA systems. In recent years, many systems exploit information published on social media to support crisis responder with a global view of what is happening. With this regards, Rogstadius et al. [153] propose a system called CrisisTracker to exploit the data gathered from social media in order to analyze large-scale events such as natural disasters. Cameron et al. [154] also propose a system (ESA-AWTM) for crisis coordination by using data gathered from Twitter. In [42], a decision support framework for improving situation awareness in disaster management via the interaction with crowdsourcing data coming from social media is proposed. Yin et al [155] describe a system architecture capable of processing text streams from Twitter to support SA during natural disasters and crises. Lastly, Thompson et al. [33] propose an architecture for identifying accidents by using data gathered by smartphone in order to support emergency responders with situation awareness of what has happened automatically.

2.2.7 Classification of the Research Focus Areas in Situation Awareness

The proposed overview of SA systems, techniques and approaches, although not exhaustive, it is useful to outline the main research focus areas of computer science in situation awareness, in order to understand which are the most common category of approaches that scholars and researchers usually adopt to overcome SA errors and demons.

Accordingly, we classified the above described works in 5 categories (as depicted in Figure 2.3) according to the nature of the core technique or approach they proposed or adopted: data mining, logic and formal theory, machine learning, computational intelligence, architecture and computing paradigm. Furthermore, considering the functional view of the SA system of Figure 2.2, it is possible to classify each of the identified research area in SA with respect to the system's functionalities in which it is mostly adopted. Figure 2.4 depicts the functional view of the SA system compared with the Endsley model. Moreover, at the bottom of

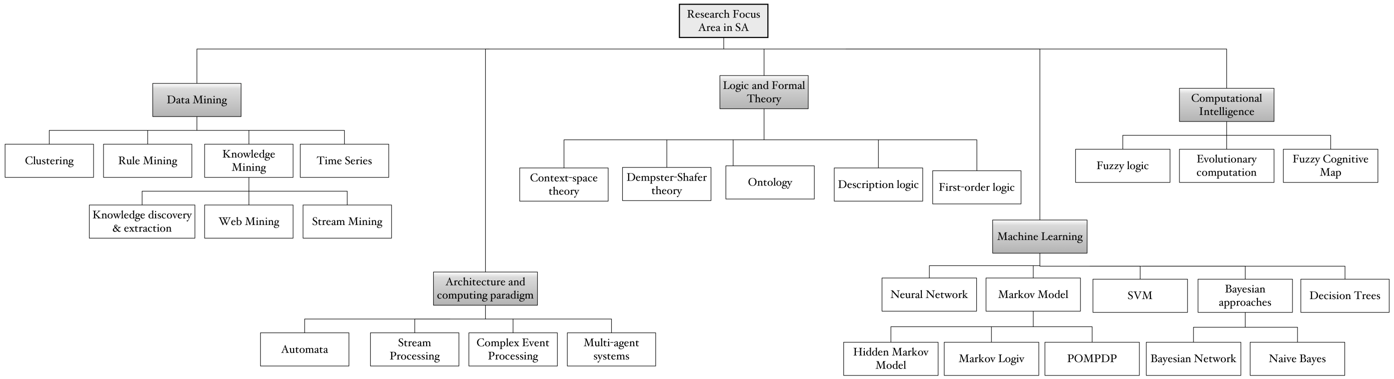


Figure 2.3 Classification of research focus areas in situation awareness

the figure, we have reported the five categories of the proposed classification of research focus areas in SA. The colored bars represent the main phases of the functional view of SA system in which the corresponding approaches and techniques are mainly used. This does not mean that a specific research focus area is used only in the indicated SA functionalities, as it would be possible to use any approach in each phase (at least theoretically). The meaning of the provided representation is that, among the analyzed works, most of them are specifically related with some phases and functionalities of an SA system. Consequently, we observe that data mining approaches are usually adopted in the initial data processing phases of the system, especially for information fusion of low-level data and for situation identification. Logic-based approaches, and especially ontology-based ones, are indeed applied for the realization of complete system for SA. Moreover, logic-based and rule-based inference approaches are widely used for supporting decision making. The approaches falling in the machine learning category are mainly exploited for situation identification. Indeed, classification approaches based on SVM, neural networks, HMM and Bayesian techniques can be really effective for the identification of situations. Moreover, the predictive capability of such approaches are also exploited for situation projection. Computational intelligence, and specifically, fuzzy logic, is used for the realization of expert-based approaches for situation identification and decision making support. Fuzzy cognitive maps are used to support comprehension of the current situation and to suggest correct decisions. Lastly, Complex Event Processing are used for situation identification and for decision making. Stream processing, intelligent agents and automata are mainly used to process data and to identify situations.

2.3 Summary

In this chapter we have introduced the theoretical background of situation awareness. Besides the various and heterogeneous defi-

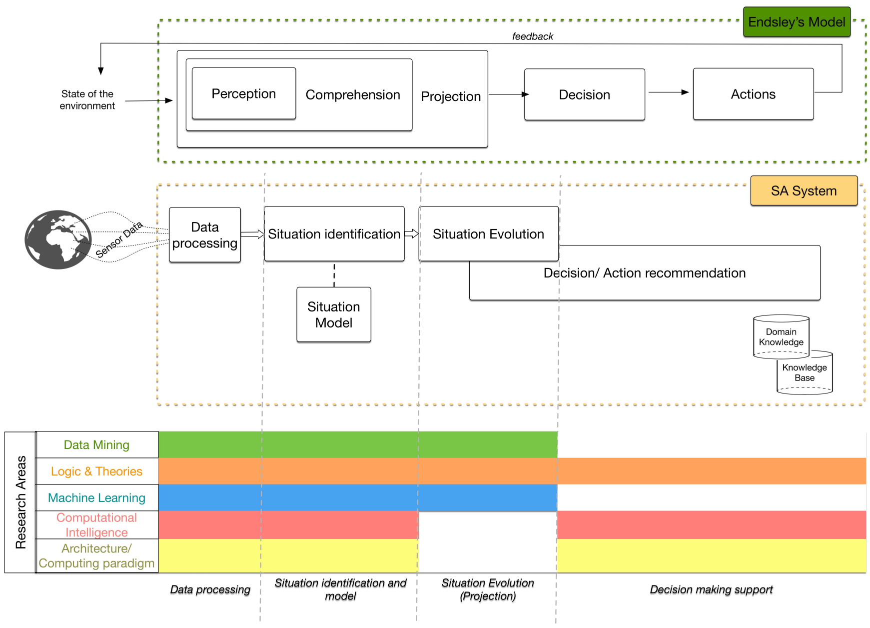


Figure 2.4 Relations between Endsley model (top), SA system (middle) research focus areas (bottom)

nitions of SA, it is straightforward that SA is a complex concept and, at the same time, a fundamental construct of dynamic decision making processes, whatever is the domain or the application. This is demonstrated by the huge number of research papers concerning the multiple aspects of the SA. Focusing our attention on techniques, approaches, models and methods adopted for the realization of an SA system, we realized an overview of the relevant literature. Such analysis allowed us to delineate the major computer science research focus areas involved in finding good solutions for advancing the support to human SA assessment and decision making. Many efforts have been put so far in this sense, leading to important results and findings and effective and useful systems capable of limiting many SA errors in specific applications. Despite that, the way to the definition of a definitive SA solution, capable to address and solve all the SA demons, is still long and winding. What clearly emerges is a great and daunting difficulty in the definition of an overall solution which can be applied, with the same effectiveness, in heterogeneous domains and applications.

In the next chapter, we analyze a novel information processing paradigm, that is the Granular Computing, in order to preliminary evaluate its applicability and usefulness for supporting SA. In the subsequent chapters, both the research focus areas analyzed in this chapter (especially data mining and ontologies), as well as the Granular Computing paradigm, will be exploited to provide an innovative support to SA by contributing to the resolution of some SA demons.

Chapter 3

Granular Computing

"Everything should be made as simple as possible, but no simpler."

— Albert Einstein

Granular Computing (GrC) is today a dynamic area of research attracting many scholars and practitioners. GrC is essentially a novel information processing paradigm focused on representing and processing basic chunks of information, namely information granules. In this chapter, we define the granular computing and provide a brief overview of the main research focus areas in GrC and of the formal frameworks used to represent information granules. Through the analysis of the theoretical foundations of GrC, it is possible to observe that this paradigm leverages on strong philosophical and methodological basis. Such theoretical foundations share many concepts and characteristics with the situation awareness paradigm. In Section 3.2 we propose our representation of an analysis about the commonalities between GrC and SA. These commonalities comfort our intuition of the possible benefits for SA deriving by the exploitation of the GrC paradigm. Furthermore, during the study of the state of the art in SA (described in chapter 2), it was found that some features of GrC techniques are used, in conjunction with other ones, for the implementation of some functionalities of an SA process with good results. However, such uses do not completely exploit the methodological and technical framework of GrC, but they limit to some specific and detailed adoption of formalisms belonging to GrC (e.g., fuzzy sets). According to our perspective, instead, a thorough exploitation of GrC (including all the methodological and technical aspects) would be desirable assuming that the use of GrC can actually improve SA processes. In order to realize a preliminary verification in terms of feasibility related to the use of GrC in SA systems, we conduct some exploratory tests and studies that lead us to propose a mapping of the main recent GrC computational techniques on the phases of the functional view of an SA system. At a first analysis, it appears feasible and useful the support of GrC to the SA at all the levels, from the perception to the performance of actions. Leveraging on the conducted researches and on our proposed contributions, which are presented in details in Chapter 7, it is possible to affirm that GrC can be seen as a valid and solid alternative to the traditional research focus areas in SA.

Parts of this chapter have been previously published in:

- Loia, V., D'Aniello, G., Gaeta, A., Orciuoli, F. (2016). Enforcing situation awareness with granular computing: a systematic overview and new perspectives. Granular Computing, 1(2), 127-143.

- Giuseppe D'Aniello, Matteo Gaeta, Marek Z. Reformat: Collective Perception in Smart Tourism Destinations with Rough Sets. 2017 3rd IEEE International Conference on Cybernetics (CYBCONF); 06/2017.

- Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Vincenzo Loia, Marek Z. Reformat: Application of Granular Computing and Three-way decisions to Analysis of Competing Hypotheses. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2016), Budapest, Hungary; 10/2016.

- Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Vincenzo Loia, Marek Z. Reformat: Collective Awareness in Smart City with Fuzzy Cognitive Maps and Fuzzy sets. 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, Canada; 07/2016.

3.1 Theoretical Background

Granular computing (GrC) [156, 157, 158, 159] is an umbrella term for a set of methods, tools and techniques that explicitly represent and process different granularities of information [160]. Bargiela and Pedrycz define the GrC as a general computation theory for effectively using granules of information such as classes, clusters, subsets, groups and intervals to build an efficient computational model for designing human-centric intelligent systems [159]. Although GrC is gaining momentum just in recent years, its roots can be found yet in the works of Zadeh [161] as a basis of computing with words.

Information granules (or simply *granules*) are "collections of entities characterized by some notions of closeness, proximity, resemblance or similarity"[162]. A granule may be considered one of the small particles forming a larger unit. For instance, a granule can be a subset of a set, a class of objects, a section of an article, or a module, a component and a service of a system. Granules can be decomposed into smaller or finer granules called *subgranules*. In order to construct or decompose granules we need to employ a two-way operation, namely *granulation*. On one hand, *construction* (or synthesizing) is related to the process of forming a larger and higher level granule from smaller and lower level subgranules. On the other hand, *decomposition* represents the basis of the process of dividing a larger granule into smaller and lower level granules.

Granules and subgranules can be organized by means of *levels*, *hierarchies* and *granular structures* as represented in Figure 3.1. Levels consist of one or more granules that are formed with respect to a particular degree of granularity. The highest level (that is the coarsest one) contains the universe of the problem; at the lowest level (that is the finest one) there are individual elements and the basic particles of the used model [163]. In the middle, different intermediate levels with different granularity exist. An intuitive example of granules and granulation is represented by *words*. Each word is a granule at the lowest level. A group of granules (i.e. of words) forms a sentence, which is a granule of the higher level. A group of sentence granules forms a paragraph granule, and so on [163]. The granular structures, composed by different levels of granulation, are used for representing and interpreting a problem or a system. Indeed, the advantage offered by a granular structure is the multilevel understanding and representation of the system or problem [164, 165, 2]. A granular structure is able to capture only a limited aspects of a problem or a system form a specific point of view. For this reason, when dealing with complex prob-

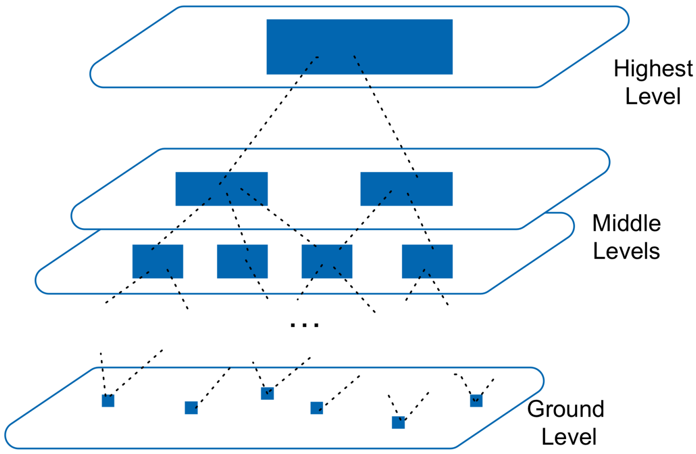


Figure 3.1 Multilevel granular structure (our re-elaboration from [2])

lem, we need a multiple view of the problem itself, by leveraging on a family of granular structure, wherein each structure gives a view for a specific purpose. Thus, multiple granular structures offer a multiview representation and understanding of the problem. Switching from a granular structure to the other and combining different views allows to gain additional insights not available in a single view.

The process of granulation relies on different kind of relationships that can be defined among granules. Yao proposes a classification of such relationships in two groups: interrelationship and intrarelationship [166]. Interrelationship is the basis of grouping small objects together to form granules based on similarity, functionality, or distance, and it is involved in the process of construction, while intrarelationship are the basis for dividing a granule in smaller granules, as happen in the decomposition process. Yao et al. in [163] outline the main kind of relationships used for granulation, which are: refinement and coarsening, partial ordering, partitions and covering, is-a relationships, similarity relationships.

3.1.1 Formal Frameworks for Granular Computing

Information granules can be represented by means of different formal settings. Yao et al. in [163] and Pedrycz et al. [167] synthesize, in their works, the most used formal frameworks.

A simple approach to represent a granule foresees the use of **sets** (interval) that can be considered as granules based on a relation of dichotomy, in which an element may belong or not to a given granule. Interval analysis introduces intervals as a means of representing real data, providing methods for numeric processing of these intervals [3]. Usually, sets construct the area of the feature space based on the high homogeneity of the patterns [168, 169].

Fuzzy sets is one of the most important conceptual generalization of sets and have been widely used to represent information granules by admitting partial membership of an element to a granule [170, 171].

Shadowed sets describe a granule by distinguishing the elements in three types: fully belonging to the granule, not belonging to the granule, belongingness is completely unknown. This allows to quantify the factor of uncertainty related to the construction of any granule [172].

Probability-based granules are represented by means of a probability density functions. The probability of each element quantifies the membership of the element to the granule.

Rough sets represents granules of elements drawn together in terms of the indiscernibility relation. Each granule is roughly described by means of its lower and upper approximation of a given rough set. Rough set is considered as one of the fundamental techniques of GrC for solving the vagueness of information, especially in data mining [3, 173].

Cluster analysis is used to build the information granules by using clustering algorithm to define the seeds of the granules which are the centers of the clusters. Different similarity measures and clustering methods can be used to this scope [174, 175, 3].

3.1.2 Justifiable Granularity

An open issue in GrC concerns finding an appropriate and systematic way of building information granules, whatever is the formalism adopted to represent granules. Given a collection of pieces of data, the problem is to form a representative information granule which reflects the nature of the available data [176]. Pedrycz et al. [176, 167] have proposed a general solution to such issue by means of an approach for designing information granules based on the principle of *justifiable granularity* which is independent from the way granules are represented. Specifically, given a set of experimental evidence $D = \{x_1, x_2, ..., x_N\}$, that for sake of simplicity we consider of numerical nature, according to the principle of justifiable granularity an information granule G should be experimentally justifiable and should exhibit a significant level of specificity. Intuitively, this means that the numeric evidence accumulated within the bounds of G has to be as high as possible to be experimentally justifiable. Indeed, the existence of the information granule is well motivated (justified) as being reflective of the existing experimental data. For instance, if G is a set (interval) then the more data are included within the bounds of G, the better is the granule as the set is more legitimate by the available data. On the other hand, the information granule should be as specific as possible. Thus, the granule should have a well-defined semantics (meaning), which implies that G is highly detailed. This implies that the smaller the information granule is, the better is the granulation.

Consequently, the principle of justifiable granularity represents a trade-off between two measures: *coverage* and *specificity*. A correct expression of these two measures depends on the nature of the set created (e.g., crisp as for k-means or fuzzy as for the fuzzy c-means) but, in general, coverage is related to the ability of covering data based on the available experimental evidence and specificity deals with the level of abstraction of the granule prototype by considering its size. As an example, for crisp sets a measure of coverage can be $Cov(G) = \frac{1}{N}\ card\{x_k | x_k \epsilon G\}$ while for fuzzy sets we can sum the degree of memberships of the elements $Cov(G) = \frac{1}{N}\sum_{k=1}^{N} \mu_G(x_k)$. Ideally, $Cov(G)$ should be 1, which means all data is covered by the prototype. Specificity requires that the intervals are as narrow (specific) as possible. The specificity of an interval can be evaluated in numerous ways. A specificity measure has to satisfy two requirements: it attains a maximal value for single-element, and the broader the interval, the lower the specificity measure. Considering that the increase in the values of coverage comes at the expense of the specificity of the information granule (and viceversa), Pedrycz et al. propose an approach to optimize the two parameters in order to create optimal (according to the principle of justifiable granularity) information granule for different theoretical frameworks [176, 167].

3.1.3 Three Perspectives of Granular Computing

The majority of existing research treats the granular computing at a concrete level, focusing on specific models and formal frameworks to represent and process data, without really considering the GrC as a new research area but more likely as an approach for information processing, focusing on a specific perspective or on a given mathematical model. Conversely, GrC can be considered, for all intents and purposes, as a new theory of computation, which Yao defines as *triarchic theory of granular computing* [177, 164, 158, 178, 2]. According to such theory, GrC is based on three pillars: a philosophy of structured granular thinking; a methodology of structured granular problem solving and a mechanism of structured granular information processing. Usually, such theory is represented as a triangle (depicted in Figure 3.2) which emphasizes the fact that each component of the theory support and, at the same time, is supported by the other two components. The powerful and expressiveness of such a theory is that it offers a comprehensive understanding of granular computing as a field of study in its own right [2], thus offering the theoretical foundations which can be used to propose new models, approaches,

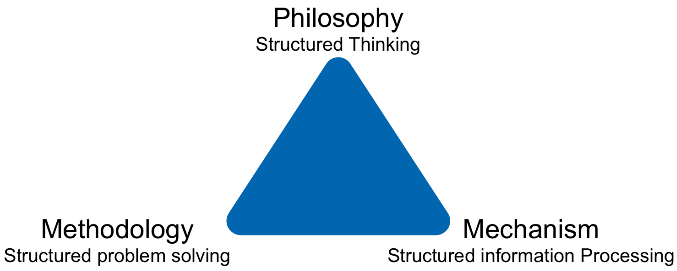


Figure 3.2 Triarchic theory of granular computing (our re-elaboration from [2])

methods as well as integrations with other research field. According to this theory, the goal of GrC is to study nature-inspired and human-inspired structured ways and approaches to thinking, problem solving and information processing [178, 2]. The purpose of GrC is to empower humans with granular computing strategies, methods and tools to make them better problem solver and to design and implement intelligent systems that adopt granular computing principles.

3.1.3.1 Philosophy of Structured Granular Thinking

The structured understanding and representation with multiple levels of granularity of a problem or a system foreseen by the granular computing leverages on solid philosophical views. First of all, the decomposition of coarser granules of the higher levels of a granular structure to finer granules follows the *reductionist thinking* approach. Indeed, the reductionism focuses on breaking a complex problem in simpler parts and on a synthesis approach for inferring properties of the whole by the properties of its parts [179, 2], which is a fundamental property for the definition of granular structures.

Another philosophical views sharing meaningful concepts with the granular computing is the *systems thinking* [180, 181]. It proposes a way of synthetic thinking for obtaining a holistic view of a system [2]. The main characteristic of this way of thinking is the ability to shift the attention of a human back and forth between the different levels of the system and to explain things in terms of their context. Quite straightforward, it is possible to identify that such different levels of the system can be seen as the multiple levels of granularity in a granular structure.

Lastly, the *levelism* [182, 183, 2] is a philosophical view in which the notion of levels is interpreted in many ways, as objective levels of reality, epistemological levels of understanding, and so on. Such ideas of levelism are directly applied to the granular computing structures, allowing for a level-wise granular processing. Thus, the above mentioned three philosophical views (reductionism, systems and levelism) are integrated and exploited in a philosophical standpoint of granular computing [2].

3.1.3.2 Methodology of Structured Problem Solving

From a methodological perspectives, granular computing means to solve problems at multiple levels by exploiting multiple granular structures. Different human-inspired problem solving approaches can be used to deal with such granular structures as top-down, bottom-up and middle-out approaches [2]. The top-down strategy is also defined as "divide and conquer". With this strategy, a problem described with larger granules is decomposed into a family of sub-problems described with smaller granules. Moving towards lower levels means to gradually add details, in order to make an abstract description more concrete. Such an approach corresponds to the analytical thinking [2]. Needless to say that the correctness of lower levels strongly depends on the correctness of the higher level, which implies that we must have a global view and a deep conceptual understanding of the whole problem for describing the problem at higher levels of granularity. The advantage of top-down strategies relies in the possibility to postpone the decisions about specific details of lower levels, allowing us to focus on the global picture without to be distracted by useless details. Unfortunately, this requires to have a good understanding of the whole problem. We can adopt the opposite approach to deal with this issue. In the bottom-up strategy, we start from lower levels of granularity and move to higher levels of abstraction, leveraging on the so-called synthetic thinking. This means that we start from the details of the problem that we have understood better and then we try to abstract them towards a global view in order to solve the whole problem. Although this approach does not require that we own a global understanding of the problem, the lack of a guidance given by an higher level can lead to a bad comprehension of the whole problem, bringing to the definition of a bad solution. A middle-out approach tries to combine both the approaches, starting from an intermediate level of detail that allows us to exploit all the detailed knowledge that we have about specific aspects of the problem, while having at least a partial view of the whole problem. A specific kind of approaches, namely the *iterative approaches* (or *hermeneutic circle*), deserves a particular attention. Indeed, such approaches try to address the issue of understanding a problem that involves both parts and a whole. The idea is that the understanding of some parts leads to the understanding of a larger whole which again can only be understood on the basis of other parts [2]. This requires to move back and forth between adjacent levels of granularity, thus by leveraging on the operations of zooming-in and zooming-out on the granular structure. Solving a problem following the methodological view of granular computing means to fully explore the relationships between all the granules of the multiple granular structures from different point of view, with no need to adopt just one of the aforementioned problem-solving approaches but by combining wisely several methods in the various stages of problem solving.

3.1.3.3 Mechanism of Structured Granular Information Processing

According to its computational perspective, GrC focuses on a paradigm for representing and processing information in a multi-level architecture also defined as an *information processing pyramid* by Bargiela and Pedrycz [184], depicted in Figure 3.3. The

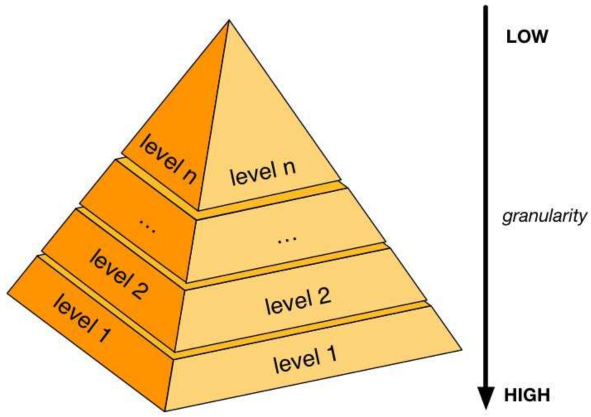


Figure 3.3 Information Pyramid

bottom level of the pyramid (that is the level where we can find the highest level of granulation in terms of number of granules, thus obtaining more finer granules) is typically associated with numeric processing. The intermediate level provides larger information granules. Lastly, the top level (usually associated with the lowest level of granulation, i.e. coarser granules) is usually devoted to symbol-based processing. The selection of the right techniques at each level is driven by the specific domain, available resources, requirements, and so on. Among these techniques, the most relevant ones associated to GrC will be reviewed and classified in next section.

3.1.4 Classification of the Research Areas in Granular Computing

Research works in granular computing address many different problems by exploiting different formalisms, approaches and strategies, thanks to the inherent flexibility of the GrC paradigm. Consequently, a classification of the most influential works according to the research areas to which they refer is helpful for understanding the main advantages, weakness, research gaps of such techniques according to the specific application we want to realize. Salehi

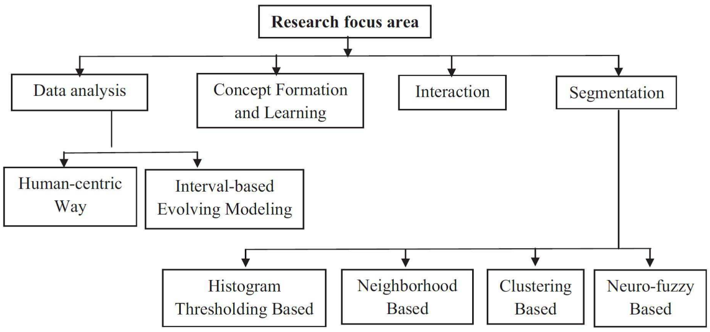


Figure 3.4 Research focus area categories in Granular Computing (from Salehi et al. [3]).

et al. [3] propose a valid classification of GrC information processing techniques and approaches along four research focus areas, as depicted in Figure 3.4: data analysis, concept formation and learning, interaction, segmentation. In what follows, we briefly describe such classification with an indication of the main characteristics of the four research focus areas identified by Salehi et al. The details of the systematic mapping study can be found in [3], with a review of the most recent contributions to GrC.

Data analysis is likely the main application of GrC. As already discussed, human-centric data analysis is fundamental in GrC, as data should be represented in an interpretable way in order to support human in the analysis process. Usually, data and relationships are defined in spatial and temporal domain through multiple granular structures [163]. Interval-based Evolving Methods is a category of approaches in data analysis used for dealing with heterogeneous stream of data in time-varying systems [163, 185].

Concept formation and learning involve the adoption of learning strategies to draw correspondences between granules (and their relationships) and concepts (and their relationships). Such correspondences is fundamental to support human interpretation of the granular structures. The proper set of concepts should be identified according to the specific problem or systems as well as the proper design strategy for learning should be considered, as different learning mechanisms lead to different descriptions of the target concepts [186, 187].

Interaction deals with discovering and modeling interactions of objects in interactive granular systems. These interactions can occur between defined objects of soft computing approaches, machine learning or data mining techniques.

Segmentation is the last category of the classification proposed by Salehi et al. [3]. Segmentation approaches are used for partitioning data (video, images, signals, text) and also for classifying such segments. The main categories of segmentation (depicted in Figure 3.4) are: histogram thresholding, clustering, neuro-fuzzy, neighborhood. Histogram thresholding is especially used in image segmentation by finding valleys and peaks in histogram [188]. In clustering based approaches the granules are formed by analyzing the proximity of the elements, using k-means, Fuzzy k-means, Fuzzy c-means and other popular clustering algorithms. In neuro-fuzzy based approaches, the granules are constructed by using the integration of processing capabilities and readability of neural networks and fuzzy rule base systems [188]. Neighborhood approaches use the criterion of uniformity among data to form the information granules via segmentation [189].

3.2 Strengthening Situation Awareness with Granular Computing

As asserted by Yao [178] as many of other scholars [190, 160, 191], granular computing can be considered as a human-oriented processing paradigm. Indeed, humans tend to organize everything in order to solve problems and make decisions, and thus the ability to classify things becomes crucial for human thinking. The results of such kind of organizations and classifications are some type of structures, especially hierarchical structures. Moreover, to have a complete understanding of a problem, humans form multiple views of the world, in order to analyze it from different perspectives, by leveraging on different representation schemes. All such characteristics of the human brain are well supported by the multiple granular structures and the multilevel/multiview approach of the GrC. Therefore, one of the main objective of GrC is the definition of new computational models that may help in the design of intelligent systems supporting human problem solving by exploiting some specific aspects and process of human intelligence [192]. This requires the definition of novel computational approaches and tools, based on the multi granular structures and multilevel views of a problem, for better supporting humans.

This objective is, indeed, very similar to the objective of SA system. Both GrC and SA aim at supporting humans in improving their performance in problem solving and decision making, respectively, by unlocking and leveraging on the mechanisms and processes of the human brain, even if with a different point of view. Indeed, granular computing is more inherent with the data analysis and with the computational aspects of information processing, while situation awareness is more focused on the cognitive and human factors of data processing and decision making. Many other commonalities are shared by the two paradigms. In Table 3.1 we present in details the results obtained by conducting a deeper analysis related to two paradigms for the identification of their commonalities, at both philosophical, methodological and technical level. The results of such analysis makes interesting and advisable the application of GrC techniques for strengthening the capabilities of SA systems in supporting dynamic decision making.

From the philosophical perspective, GrC and SA share a similar approach to the problem solving and decision making based on the cognitive mechanisms of human brain, exploiting the capability of structured thinking. From a methodological perspective, GrC and SA share the same approaches to problem solving and information processing, and similar concepts related to multiview and multilevel analysis (see Table 3.1 for further details). Lastly, considering the mechanisms of information processing, both the paradigms focus on the concept of information pyramid in order to define different levels of abstraction of data and to exploit different representation schemes and processing techniques at each level.

Table 3.1: Commonalities between GrC and SA

View	ID	Granular Computing	Situation Awareness
Philosophy	1	Focus on human-inspired problem solving	Focus on human dynamic decision making
	2	Understanding the underlying principles and mechanisms of human problem solving and human structured thinking	Understanding the cognitive mechanisms that influences and drives the decision making processes
Methodology of problem solving	3	Different approaches to problem solving and information processing: top-down, bottom-up, middle-out, etc.	Combining different approaches for processing information (e.g., the alternation between top-down (goal-driven) and bottom-up (data-driven)) is crucial to SA to address data overload and attentional tunneling daemons
	4	Multiple views of the granular structures support the analysis from different points of view and multiple descriptions of problems	Multiple goals allows the human operator to focus on the relevant details from different perspectives
	5	Multiple levels of granulation allows to describe elements with different details and with different representation schemes	Focusing only on the relevant details (using mental models) is an important capability to maintain high level of SA
	6	The principles of focused efforts is applied to granular structures to recall the attention on the focal point	The goal drives the attention of human operator to specific information processed according to a given mental model

Continued on next page

Table 3.1 – *Continued from previous page*

View	ID	Granular Computing	Situation Awareness
	7	Decomposition of complex problems and structures in simpler ones	Decomposition of goals in subgoals and SA requirements (e.g., GDTA approach). Situations are composed of simpler elements (contexts, objects, relations)
Mechanism of information processing	8	Granular structures, consisting of different levels and relationships, give a specific representation of a problem (as depicted in the information pyramid of Figure 3.3)	The concept of information pyramid is fundamental in SA to represent the abstraction of low-level data in higher levels concept relevant for the domain (Figure 3.5)
	9	One goal of GrC is the design of intelligent systems based on the mechanisms of human brain	SA systems are intelligent systems which try to mimic and support the cognitive processes to increase SA
	10	Dealing with uncertainty and vagueness of information	SA systems have to address issues related with reliability of sensor data characterized by uncertainty
	11	Representation of the granules and granular structures in a formal and precise way is a basic issue in GrC	Representation and modeling of situations (also in a formal way) is a major issue in designing SA systems
	12	The process of granulation is a fundamental step in GrC as it allows for the construction of granules, levels and hierarchies	The process of situation identification is a fundamental step in SA systems since it is the basis for supporting comprehension, projection and decision making

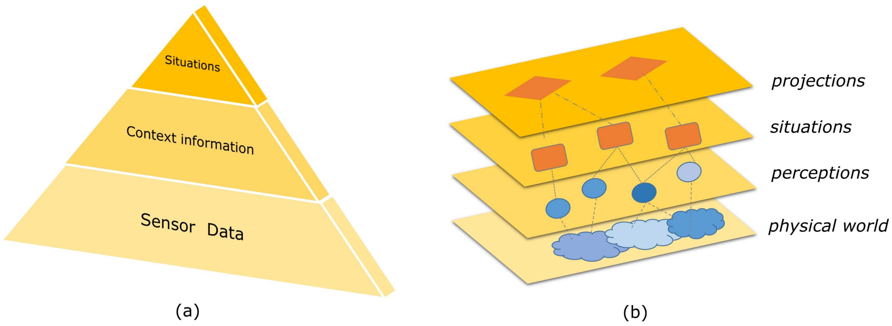


Figure 3.5 Information Pyramid in Situation Awareness

In Section 3.1.3.2, we described the information pyramid in GrC for the processing of information at different levels of granularity. A similar information pyramid is also widely used in situation awareness, especially in situation identification tasks, as sketched in Figure 3.5.(a). For instance, the authors of [193] provide an approach to infer knowledge on situations in a physical environment, equipped with sensors, by exploiting three levels of computation. At the lower level of the pyramid, sensors (light sensors, microphones, GPS, biosensors, body temperature, etc.) and logical sensors (time of the day, schedule of the day, universal or known facts, etc.) observe features of the world and provide data. Sensors data become inputs for the intermediate level where context information are deduced and positioned. Here, context is considered as any information about user, his/her environment or activities. At the top level, a number of context information, occurring within same time frame, are transformed in abstract actionable objects, namely situations. This hierarchy of data organization increases the usefulness of data and decreases its size. The concept of information pyramid is also considered in [194] where the authors introduce four planes of abstraction by considering both the Endsley's Model [25] and the JDL Data Fusion Model [195], as depicted in Figure 3.5.(b). The bottom plane is the physical world whose aspects can be monitored. The next plane is that of perception where we can find the representation of objects of the physical world that are observed through sensors. The next plane represents situations, i.e., the knowledge about the objects in a specific area. At the top plane, we can find projections, which are symbolic information to anticipate future events and their implications. Beyond the concept of information pyramid, at the processing level GrC and SA share other concepts, like the need to deal with uncertain and vague information and the need of formally and explicitly represent data and complex structures, as reported in Table 3.1.

3.2.1 Granular Computing supporting Information Processing in Situation Awareness

In Table 3.1 we outlined the main commonalities between GrC and SA paradigms according to the three views of GrC. Such analysis would seem to motivate, at least from an abstract point of view, a synergy of the granular computing techniques in situation awareness systems. Such a synergy is made more evident from the results of a preliminary feasibility analysis of the specific GrC computational techniques for supporting SA, which we describe in what follows.

We realize a mapping between the main granular information processing techniques and the information processing phases of the functional view of an SA system, according to the Endsley model. The Figure 3.6 depicts our mapping between GrC techniques and the phase of Endsley model. Following such mapping, we propose an overview of GrC techniques to assess their usefulness in the context of SA.

As shown in Figure 3.6, the Data Analysis techniques are useful for all the phases of SA (perception, comprehension and projection, since they naturally support the requirements of information representation needed at all the levels of the information pyramid. The segmentation approaches are mainly compliant with the low level of the SA information pyramid for supporting the perception phase, while concept formation and interaction techniques are enabling techniques for the comprehension and projection phases in

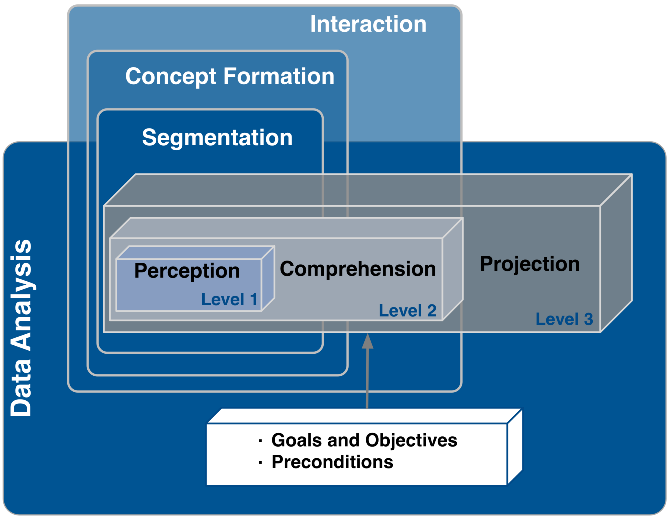


Figure 3.6 GrC techniques and SA

terms of feasibility.

Most of the methods and techniques analyzed in our overview can support the first level of SA in addressing issues related to a proper organization of the information around goals and objectives, i.e. building information granules, and to object assessment and state estimation via techniques of outlier detection, attribute reduction and data conflict resolution. For the level 2 SA, we focused our attention on techniques that can support in understanding the right meaning of the elements perceived and, thus, mainly techniques of concept formation. For the level 3 SA, the analysis is focused on techniques supporting prediction in a future state of some elements of a recognized situation. In what follows, we briefly report an overview of these GrC techniques according to the three levels of SA.

3.2.1.1 Perception

The perception level of the SA model is devoted to perceive and recognize elements of the environment by combining observations and measurements from different sources. At least two main issues related to the perception level can be solved with GrC: designing information granules around goals and support information filtering via analysis of outlier, conflicting and spurious data and attribute reduction.

The principle of justifiable granularity [196] allows the construction of granular descriptors according to linguistic characterization. Fuzzy clustering, in particular Fuzzy C-Means, is employed as a vehicle to build information granules. The problem of successive refinement and generalization of prototype information granules is discussed in [197]. Construction of information granules according to the principle of justifiable granulation fits well with the SA requirements of selecting the proper information and at the right level of abstraction for the specific goal and objective. These approaches work well when there is the availability of experimental evidence resulting from previous and similar situations that can be stored in schemata and scripts. Moreover, the principle of justifiable granularity is applicable also to time series [190] and, in a former proposal, applied to signal analysis [198]. With regard to signal processing, another interesting perspective is the hybrid method based on neural networks, GrC and evolutionary computing proposed in [199].

Recently, Sanchez et al. [200] have proposed an approach for constructing information granules based on the theory of the uncertainty that can be useful in environments characterized by high level of uncertainty and noise, such as sensor networks. The basic idea employed in the work is that a reduction of uncertainty can be obtain by the difference of two uncertain models of the same information, e.g. *a priori* and *a posteriori* models.

A number of GrC techniques can be employed to support recognition of elements, information filtering and attribute reduction requirements of SA. Some recent methods supporting recognition of the elements of an environment are employed with Granular Neural Network. It is the case of the method proposed in [201] that has been tested with human recognition based on the face biometric measure. The method is based on the adoption of a modular neural networks optimized with a hierarchical genetic algorithm, and GrC is used to split the whole database into sub modules. Granular neural networks have been employed also for classification of land use/cove images [202], and for fusions of numeric and linguistic data [203], and this last case appears of interest in several SA scenarios where important sources of information can be textual (e.g. social media).

With regard to filtering the most relevant data for fusion and classification objectives, the work of [204] on the adoption of rough integrals in order to select the most informative sensors for a specific objective can be used. A similar approach is employed in [205], and an hybrid approach combining fuzzy and rough set for classification under uncertainty is presented in [206].

Analysis and detection of outlier and spurious data is investigated in several works such as [207] [208] [209] [210] [211]. A commonality among these works on outlier detection is the adoption of Pawlak theory of rough sets and its capability of approximating sets to detect outlier objects having abnormal attributes and properties (generally in boundary regions). In [209] is used the concept of Non-Reduct to discover a set of attributes that may contain outliers, [210] proposes the adoption of outlier detection algorithm based on the neighborhood rough set model, [211] introduce the concept of GR-based outliers and proposes a detection algorithm working on this concept. With regards to spatiotemporal requirements, a specific application for detecting spatial and temporal outliers is proposed in [212].

Attribute reduction plays a key role in applications requiring SA since high dimension data are common and this requires high computational time and space. Several works propose rough sets and GrC methods to solve the attribute reduction problem. An application of rough set theory for attribute reduction to support situation recognition via classification of precursory information in reference to earthquake rupture analysis is proposed in [213]. Recently, [214] proposes a generalized framework allowing human expert to specify conditions in terms of group of measures and thresholds which are relevant to user requirements or real applications. The proposed framework gives the possibility of choosing the appropriate reducts on the basis of users and application requirements, and this can support the goal-oriented information processing principles of SA based design.

An issue that arises in concrete applications demanding SA is resolving data conflict, e.g. when we have multiple values of an observation or a variable that are not compatible. In [215] it is presented a multi-sensor data fusion framework including voting-like process to resolve conflict among data using a measure of compatibility. An alternative approach to the voting process can be the adoption of soft-consensus model supporting human-like perception processes [216].

3.2.1.2 Comprehension

The second level of the Endsley model is devoted at understanding what data and cues perceived in the first level mean with respect to goals and objectives. Comprehension is achieved via a meaningful integration and prioritization of the elements of the environment perceived in the level 1 SA. GrC for concept formation supports comprehension presenting information required to this level. Based on the triarchic theory of GrC, the approach described in [163] proposes two strategies for concept learning, namely, an attribute-oriented strategy for searching a space of partitions and an attribute-value oriented strategy for search space of coverings. A perspective focused on cognition of concept learning via GrC is analyzed in [217] and [218], and proposals for building tools for automatic understanding of data via granular cognitive maps are presented in [219]. Other approaches leverage on Formal Concept Analysis (FCA) and concept lattice. It is the case of [220], where the authors propose an algorithm for generating interval-valued fuzzy formal concepts using the properties of interval-valued fuzzy graph and Galois connection and incorporation of interval-valued fuzzy graph to the concept lattice. The authors in [221] address the issue of knowledge reduction in formal concept analysis via the adoption of GrC and, specifically, via the concept of granular reduct of a formal context (i.e. a minimal attribute set preserving the object granules of a concept lattice obtained from a full attribute set). These approaches can be useful if formal contexts of the situation to recognize are available in schemata or scripts.

In [222] and [223] is defined the approach of Interactive Rough Granule Computation as a way for modeling interactive computation with rough set and other soft computing approaches. The concept of Complex Granule is defined, which allows to link traditional information granules to physical objects, and can be used by agents to make decision via adaptive (intuitive or rational) judgment. By means of interactive hierarchies of complex granules, authors evidences how it is possible to approximate vague and complex concepts that can be relevant in processes of situation recognition and decision making, such as *safe driving* in traffic control applications.

3.2.1.3 Projection

This level is devoted to project in the near future elements of the situation recognized in the previous level. Some models useful for the projection step of SA include time series and regression-based [224]. Description and prediction of time series has been deeply investigated in GrC, and we mention just few recent works. The authors in [225] propose a framework in which information granules are based on time windows, amplitude and change of amplitude, and employ fuzzy relations to predict amplitude and its change. In [226], the issue of long-term prediction is addressed via the development of a forecasting model combining a modified fuzzy c-means and information granulation. In [227], authors employ fuzzy cognitive maps to describe granular time series (built with fuzzy c-means clustering algorithm) and perform predictions.

3.2.2 Granular Computing as a Research Focus Area of Situation Awareness

In previous section we observed that the GrC computational techniques, belonging to the different research focus areas proposed in

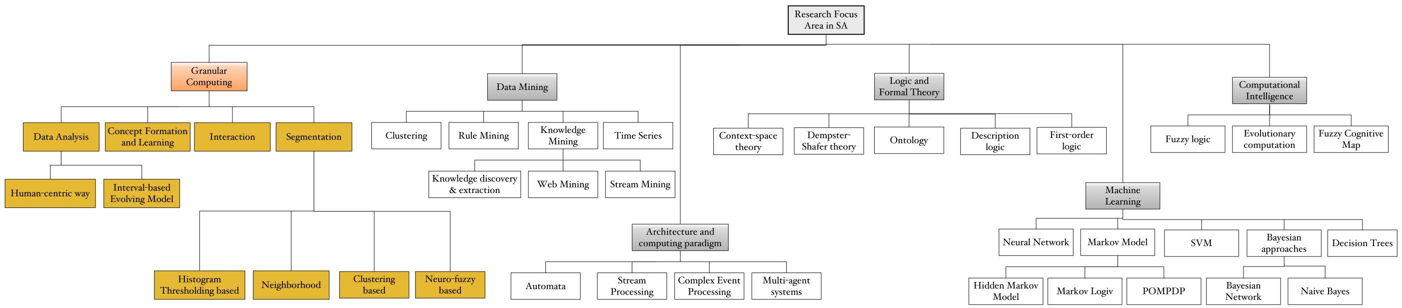


Figure 3.7 Classification of research focus areas in situation awareness including Granular Computing

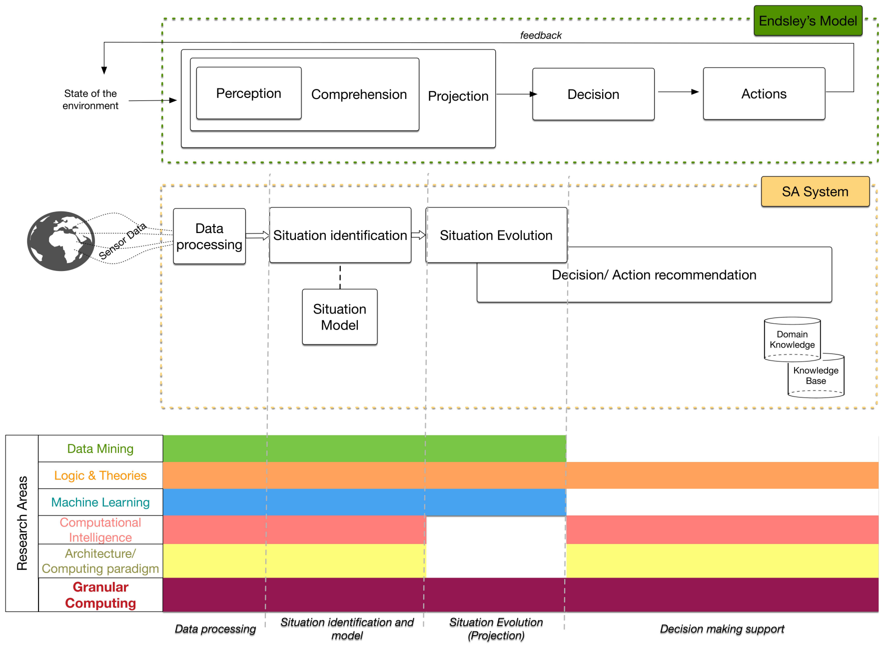


Figure 3.8 Relations between Endsley model (top), SA system (middle) research focus areas (bottom) including Granular Computing

[3], can effectively support all the three levels of SA. Thus, the use of GrC in SA systems may lead to different benefits for improving SA of human decision makers. Following such considerations, our vision is to consider the GrC as a valid, alternative research focus area in situation awareness. Accordingly, we extend the classification of research focus areas in SA (proposed in Chapter 2) with the introduction of this new research focus area, i.e., the Granular Computing, as depicted in Figure 3.7. Moreover, we redefine the mapping of the research areas with the functional view of an SA system, as initially depicted in Figure 2.4, now by considering also the GrC. The new mapping, depicted in Figure 3.8, considers the GrC as a paradigm that can be exploited for supporting all the phases of the SA and decision making.

3.3 Summary

In this chapter, we have introduced the theoretical background of GrC. GrC is an umbrella term for a set of methods, tools and techniques that explicitly represent and process different granularities of information, by means of different formalisms. We have discussed that the GrC as a new theory of computation which is grounded on solid basis from a philosophical, methodological and computational perspective. By analyzing the state-of-the-art techniques and approaches of GrC and by conducting some preliminary verifications and studies about such techniques, we noticed that GrC can support all the three levels of SA effectively and that GrC and SA share many meaningful concepts and principles. Leveraging on the major outcomes of this analysis, in Chapter 7 we define a theoretical framework based on GrC techniques and model for supporting approximate reasoning in SA and for representing situations.

Chapter 4

Research Contributions and Evaluation

"Research is to see what everybody has seen and think what nobody has thought."

— Albert Szent-Györgyi

In Chapters 2 and 3, together with the overview of SA and GrC technique, some initial contributions of this thesis have been already presented, and specifically: i) an original overview of SA techniques, models, approaches and systems with a mapping between the research focus areas to which they refer and a functional view of a generic SA system in Chapter 2; ii) a mapping between GrC techniques and the functional view of an SA system, to show the usefulness of GrC at all the levels of SA in Chapter 3. In this chapter, related to our main research objective that is to contribute to the resolution of SA demons by means of novel computational approaches and models, we outline the main research contributions of the thesis with an indication of the SA errors they address. Furthermore, in this chapter, we discusses some methodological issues related to the evaluation of Situation Awareness, thus describing qualitative and quantitative methods we adopt to evaluate the proposed research contributions. The succeeding chapters (5 to 7) present in details the proposed contributions together with the results obtained by applying the evaluation methods presented herein.

4.1 Research Contributions

Despite all the efforts lavished for addressing SA demons by means of the design of approaches, techniques, models and systems, many great research challenges are still far from being resolved, with many specific issues hindering the process of SA assessment in heterogenous domains and applications. Our objective is to give some contributions to the research area of the Situation Awareness by addressing specific problems arising at the different levels of SA, from perception to decision making. Briefly, our research contributions consist of computational models, approaches and techniques, inspired by cognitive models and approaches, able to mitigate specific issues in SA, as we describe in what follows.

In Table 4.1 we summarize the proposed contributions. On the rows, there are the main SA errors as identified by Endsley [14], and briefly described in chapter 2, grouped according to the SA level. On the columns there are our main research contributions. Notice that the contributions have been grouped by the preeminent SA research area that is used, as reported on the first row of the table. An "X" in a cell means that the proposed approach or technique gives a contribution to the resolution of the corresponding issue.

Specifically, as we evidenced in chapter 2 and 3, one of the main issues at the level 1 SA relies in data reliability: sensor data may be uncertain, missing, difficult to observe, causing serious issues on the whole SA assessment process. This issue, besides the characteristics of the physical sensors and of the sensor network, is affected also by the presence of SA demons like data overload and complexity. Our contribution in this sense is a data imputation technique based on association rule mining which is able to replace missing or low-quality data. This technique is integrated

Table 4.1 Overview of the research contributions

Research Area	**Rule Mining**	**Ontology**			**Granular Computing**	
Contributions	Quality-aware sensor data management	Adaptive Goal-driven Situation Management			Granular Computing for Situation Awareness	
Approaches / *SA Errors*	**Quality-aware sensor management via association rule mining**	**Hybrid goal-driven and data-driven approach**	**Ontological approach for goal-driven situation management**	**Adaptive goal selection approach**	**Granular computing framework for approximate reasoning**	**Situation model based on Rough Sets**
Level 1						
Data not available	x					
Data difficult to perceive						x
Failure to observe data						x
Omission	x					
Attentional narrowing		x	x	x		
High workload			x	x		
Misperception of data					x	x
Memory Failure						
Level 2						
Poor mental model				x	x	x
Wrong mental model		x	x	x	x	
Over-reliance on default values					x	
Memory failure						
Level 3						
Poor or wrong mental model				x	x	x
Other causal factors						
Maintain multiple goals		x	x	x		
Habitual schema		x				
Evalaution Method	Evaluation of the accuracy using a real dataset	SAGAT evalautions + numerical simulation			Case studies	
Chapter	*Chapter 5*	*Chapter 6*			*Chapter 7*	

in a semantic-based approach for the quality-aware management of sensor data which allows the users to virtualize physical sensors, thus to avoid dealing with issues related with the management of sensors at a physical level. The approach is ‘quality-aware’ in the sense that it allows the users to define their own quality requirements, by defining different virtual sensors, and the approach will try to provide the users with data compliant with such requirements by exploiting the data imputation technique in case of low-quality or missing data. The aim of the approach is to address errors at level 1 SA, specifically for dealing with the unavailability of data and human failures due to the difficulty in perceiving and observing data. The approach is described in chapter 5.

The other open-issues in SA that we considered are related with the SA demons of data overload, attentional tunneling, misplaced salience, and out-of-the-loop demons, which negatively impact on the capability of perceiving information and making coherent decisions. To avoid such issues, Endsley stresses the importance of alternation between goal-driven and data-driven information processing to avoid attentional tunneling and data overload problems. Accordingly to such principle, we propose an ontological approach for cognitive situation management, namely the Adaptive Goal-driven Situation Management, which supports the human operators in switching between goal-driven and data-driven information processing. Moreover, via an adaptive goal selection mechanism, it helps human operators in considering the most important goal at a given time. The most suitable goal, identified by means of desirability functions, is suggested to the user, thus concretely avoiding both attentional tunneling and data overload. The approach adapts the process of goal selection to the preferences of each user by exploiting a reinforcement learning technique. Overall, the Adaptive Goal-driven Situation Management approach addresses different errors at all the three levels of SA, as indicated in Figure 4.1: it helps in avoiding attentional narrowing and reducing the workload at level 1, it supports the use of the correct mental model at level 2 and 3, and most of all, it supports the human operators in dealing with multiple goals. The approach is described in Chapter 6.

Lastly, by leveraging on the mapping of GrC and SA we proposed in previous chapter, we further demonstrate the benefits arising from the application of GrC in SA by means of the definition of a set-theoretical framework for supporting approximate reasoning in situation awareness. This approach deals with the SA demons of attentional tunneling, data overload, complexity, wrong mental models, by addressing errors at all the three levels of SA, and specifically related with the misperception of data (level 1 SA), the use of wrong or poor mental models, and the use of false expectations (level 2 and 3 SA). The proposed framework guarantees a high degree of flexibility in the process of granulation allowing to satisfy the wide variety of requirements for perception and comprehension of situations where some elements must be perceived per similarity, others per spatial proximity, some must be fused to improve their comprehension, and so on. Moreover, it supports the approximate reasoning in situation awareness via a granular structure representing a snapshot of a situation. The granular structure is a building block for the development of tools and techniques to reason on situations in order to reduce situation awareness errors and accelerate the process of decision-making. To this purpose, we propose a technique to support operators in the analysis of conformity between a recognized situation and an expected one. Moreover, by representing situations with granular structures we can support operators in having rapid and indicative measures of how two situations, e.g. a recognized and a projected one, may differ. Leveraging on this set-theoretical framework, we define a new formal and interactive model for representing situations based on the Rough Sets theory, which is able to improve comprehension of situations and support reasoning on projections of situations. The granular computing approach for situation awareness and the model of situations based on Rough Sets are described in chapter 7.

4.2 Evaluation

In this section, we propose a brief overview about the techniques for SA measurement and we motivate the choice of using SAGAT to evaluate the proposed Adaptive Goal-driven Situation Management approach. Next, we provide further details on SAGAT and its application in an evaluation context. Lastly, we outline the methods we adopt in this thesis to evaluate all the proposed research contributions.

4.2.1 Theoretical Background on Situation Awareness Measurement

Evaluating and measuring SA is not a trivial task. An intense and heated debate about the measurement of SA is opened since a long time in the scientific community. Some scholars, although they recognize the important value of SA to operators, asserted that the utility of SA is limited because we rarely can measure it. On the other hand, according to Endsley and many other researchers, the opposite is true, and the fact that measuring SA is difficult does not mean that we can not evaluate it with scientific and rigorous approaches [228]. Indeed, the main difficulty is to effectively measure indexes related to SA instead of other factors that are involved in the task (e.g., workload, attention), and that only partially contribute to the SA. Definitely, the important point is to be able to identify what we want to measure and how. One can not simply measure the total amount of SA of an operator and try to postulate some effects on the decision making performance, as this may not be of great validity and usefulness. One can instead measure and quantify which is the improvement in the levels of SA following the implementation and definition of a new approach or system's feature in an already existing system. Indeed, the objective of SA evaluation is to measure how much the design of new approaches, methods and systems contribute (positively or negatively) to the SA. Briefly, this means that we need to think of measuring SA in relative terms: we should evaluate specific design concepts and approaches with relative comparisons. In this way, we can measure the relative improvement of SA given by the definition of a novel approach or technique with respect to an existing one or to a previous design. Lastly, one should consider that there is not too much SA neither a minimum, sufficient level of SA to achieve. Higher level of SA is always better, since this means that the operator has a better knowledge of what is happening and of what will happen next with respect to the active goal, and thus the more is the SA, the higher is the probability of performing well a given task.

Having asserted the importance of evaluating the SA and the fact that SA should be measured in relative terms, the issue is how we can effectively measure it. Numerous approaches for measuring SA have been proposed so far. Indeed, an evaluation approach for SA should not just propose a performance index reflecting the level of SA, but should also define the measurement context, the way by which an eventual simulation should be performed, the constraints on the testing situation, given that these aspects have a considerable impact on the reliability and veracity of the measure. In literature, the SA measurement methods are usually divided in the following four categories [229]:

1. *Direct system performance measures:* using evaluation scenarios specifically designed to evaluate system performance measures (e.g., time required to detect an anomaly). Usually, this foresees the introduction of disruptions to disorient the operators, or the introduction of erroneous data to measure the capability of the operators in detecting anomalies.

2. *Direct experimental technique:* using techniques based on queries or probes and measures of information seeking. This represents the most common measurement method. The Situation Awareness Global Assessment Technique (SAGAT) is one of the most widely used technique [19]. It foresees the suspension of the simulation and to ask random questions about the state of the task before resuming the action. Many other techniques have been derived by SAGAT. The differ-

ence between the techniques belonging to this class and the direct system performance measures is that the former allow measuring the SA globally, while the latter allow measuring only a specific aspect of SA through performance assessment. A specific group of direct experimental techniques uses the eye movements and eye-blink response as direct information seeking measures.

3. *Verbal protocols:* using information recorded from the observer during or after a simulation, an exercise, a video replay of a specific situation. Usually, this technique is used in the first stage of the evaluation and of the design of a system.

4. *Subjective measures:* using self-assessments, expert judgements, peer ratings, supervisor ratings, to provide subjective measures of the SA when objective measures (like in the direct experimental techniques) are difficult to obtain. A widely used approach is the Situation Awareness Rating Technique (SART) [230], that consists of an equation that produce an overall estimate of the subject's SA by means of the combination of a set of sub-scales for obtaining an integrated measure of SA. One issue with this approach is that it tends to confuse SA with workload. Generally, the issue with this category is that the measures only reflect the self-awareness of the operator, usually providing a measure of the confidence rating or preference of the operator, and thus they should be used to measure such aspects rather than the SA.

The choice of using a specific technique should be made with respect to the practical situations in which they should be used, considering their relevance and usefulness. In literature, some considerations are reported about what measures are good for what purposes [228]. As above described, the verbal protocols are best for a preliminary evaluation in order to understand SA requirements during the design phases or to establish which could be the direct objective measure to use for the specific application. Subjective measures are used when no objective quantitative data can be collected, both for the costs or the time it requires, or because it is simply impractical. Direct system performance measures obtained by scenario manipulations should be used for evaluating individuals' ability to meet some SA requirements of a specific scenario, rather than the capability of an approach or a system's feature of improving the overall SA for any individual. When possible, objective measures of SA should be obtained by means of direct experimental technique. For our purposes, in order to evaluate the improvements of SA related to the implementation of the Adaptive Goal-driven Situation Management approach, we adopt SAGAT. The use of SAGAT provide us with the following benefits:

- an objective measurement of SA for all the levels that are sufficiently independent by the specific characteristics of the subjects involved in the experimentation scenarios;
- the reliability, sensitivity, and validity of SAGAT has been experimentally proved in many studies [231];
- it is useful when the aim is to evaluate the overall SA rather than evaluating a single, specific aspect of a scenario, and this is our case;
- it supports in the evaluation of relative comparisons between different versions of a system or system's features, thus helping the designer in the realization of better systems supporting SA.

A big drawback of SAGAT is that it requires real systems/applications (at least prototypical versions) for executing the evaluation and a quite big effort for the definition of the queries related to the SA requirements. So the use of SAGAT is quite costly and time-consuming, but in turn, we obtain a reliable, replicable, worthwhile evaluation of the SA. In the next subsection, we provide some useful details about SAGAT and its use.

4.2.1.1 Details on the Situation Awareness Global Assessment Technique

The Situation Awareness Global Assessment Technique (SAGAT) assesses the level of operator SA requirements, allowing for the measurement of the global SA across all the levels (perception, comprehension and projection). SAGAT is based on a simulation employing a system of interest that is frozen at randomly selected times; operators are queried about their perceptions of the situations at that random time. In these selected intervals, the simulation is suspended while subjects answer the questions in a given time window (whose length is randomly identified). It is an objective measure of the SA since the queries allow for collecting detailed information about subject SA that can be compared and evaluated against reality. Moreover, the level of SA is directly measured via the operator's perceptions rather than inferred from behaviors that may be influenced by many other factors, as instead happens in subjective assessment involving external observers.

The execution of an evaluation with SAGAT requires the availability of a real system that can be used by the subjects during the simulation. The evaluation requires the definition of queries that should be relevant with respect to the SA requirements of the operators. Thus, they are usually defined according to a list of SA requirements that have been identified by using a form of cognitive task analysis called goal-directed task analysis (GDTA) [15]. The advantage of using GDTA is to identify the major goals of a particular task and its operational subgoals. Then, associated with each subgoals, the major decisions that the operator should made are identified. This allows for identifying the SA requirements (at all the three levels) that are needed to make these decisions and thus satisfy the related goals. Focusing on some of the subgoals that are object of the evaluation (i.e., the ones that involve the systems' feature that we want to evaluate), we define the queries for probing all the SA requirements of such goals. The queries should be formulated in a way that they require the minimum effort to the subjects of the evaluation to be understood:

they should be written using the adequate jargon of the domain, they should be easily to be responded (e.g., by using a graphical interface and multiple answers), they should be meaningful for the simulated scenario.

The procedure for executing an evaluation with SAGAT should be defined according to the following recommendations, provided by Endsley in [231]:

- The subjects involved in the evaluation should be trained about the SAGAT procedures before the testing phase.
- The subjects should attend their tasks as they normally would. Answering SAGAT queries should be intended as a secondary task.
- No displays and visual aids should be visible while subjects are answering to queries.
- No penalty should be used for guessing, since the subjects should be encouraged in answering all queries. The fact that a subject is not able to answer a query can be meaningful from the point of view of the SA.
- Subjects should not talk and exchange information among them.
- Queries submitted during a freeze of the simulation should be randomly selected from a constant set of queries.
- Some important queries can be presented at every stop; other can be omitted if they are not meaningful for the scenario. It is always important that some queries are randomly submitted to the subjects.
- The timing of each freeze should be randomly determined, thus subjects can not prepare for it.
- Usually, no freezes should occur earlier than 3 minutes into a trial.

- Multiple freezes can happen in a trial. Three stops in a 15 minute trial can be a good rule of thumb.
- A freeze should last until a certain amount of time has elapsed (from 2 minutes to 5-6 minutes), regardless of the queries.
- The number of trials necessary will depend on the variability present in the scenario and the number of data samples taken during a trial. Usually, between 30 and 60 samplings per SA query (across subjects and trials) with each design option are adequate.
- The answer to each query should be evaluated as correct or incorrect based upon if it falls into a tolerance band around the actual value (contained into a ground truth of the queries).

4.2.2 Evaluation Methods

The three main contributions outlined in Section 4.1 provide different techniques and approaches to improve, from different perspectives and with respect the different SA demons, the SA of the human operators. Considering the different nature and characteristics of the contributions, their evaluation should be conducted by using the techniques that best fit with them. In what follows, we describe the evaluation methods we adopt to evaluate the research contributions. The evaluation methods we used for each contributions are also reported in Table 4.1.

First of all, the quality-aware sensor data management approach aims at improving the reliability of sensor data by means of a data imputation technique based on association rule mining. Thus, in this case, to evaluate the capability of the technique to provide the users with correct data, we apply the approach on a real dataset containing temperature and humidity readings of real sensors, i.e., the Intel Berkley Lab Dataset [4]. By using this dataset, we perform experimentations to measure the accuracy in the reconstruction of missing data by evaluating the root mean square error (RMSE). We compare the RMSE of the proposed approach (at different rates of missing data in the dataset) with the ones of other well-known data imputation techniques: average imputations, kNN, k-means, SVD. The results of the evaluation are described in Chapter 5.

Regarding the Adaptive Goal-driven Situation Management approach, as already anticipated, we use the SAGAT approach to evaluate its impact in the improvement of the SA. Specifically, we implement the approach in three different prototypical systems: i) a dashboard for sustaining learners in self-regulated learning, in which the approach helps learners in making decisions regarding their learning activities and objectives; ii) a green fleet management system, in which we use the approach to sustain logistic operators in the management of a fleet of vehicles; iii) a decision support system (DSS) for the management of logistic operations in port container terminal for containers handling.

Furthermore, we evaluate the effects that the improvements in term of SA have on the performance of decisions, by realizing a numerical simulation of a real process regarding the management of logistic operations in a port container terminal. Specifically, we realize a discrete-event simulation in Arena of the port container terminal of Salerno, in order to measure some Key Performance Indicators (KPIs) (e.g., time needed for unloading a container, throughput) related with the logistic operations given a specific workload. Then, we measure the improvement of this KPIs when implementing the Adaptive Goal-driven Situation Management approach in the simulation.

Details about the evaluation with SAGAT and the numerical simulation, together with the obtained results, are in Chapter 6

Lastly, regarding the GrC theoretical framework for sustaining SA, we preliminary evaluate its capability of sending early warnings to human operators. The conformity analysis technique that is based on the proposed theoretical framework is a means for confirming the expectations of human operators. Therefore, we evaluate the usefulness of conformity analysis technique as a way to quantify the differences between the human expectations about situations and the actual ones. Moreover, we demonstrate the capabilities of supporting perception, comprehension and projection of the model of situation based on rough sets by means of a detailed case study regarding the movement of vessels near a harbor, in order to identify and anticipate potential dangerous situations. The details about these case studies are in Chapter 7.

4.3 Summary

In this chapter, we briefly outlined the main research contributions of this thesis: an approach for the quality-aware sensor data management implementing a data imputation technique; an adaptive goal-driven situation management approach; and a set-theoretical framework of GrC for SA, with a situation model based on rough sets. Such computational approaches, models and techniques aim at contributing to the resolution of the SA demons by addressing the most common SA errors at all the three levels of SA. Considering that the nature of the proposed approaches are different, since they aim at solving different issues at different levels, their evaluation needs to be executed with different and proper methods. Accordingly, in this chapter we outlined the evaluation methods we adopt in this thesis, motivating the choices of the proposed methods.

In the next chapters, from 5 to 7, we describe the main contributions of this thesis together with the evaluation results.

Chapter 5

Quality-aware Sensor Data Management in Situation Awareness

"The errors which arise from the absence of facts are far more numerous and more durable than those which result from unsound reasoning respecting true data."

— Charles Babbage, *On the Economy of Machinery and Manufactures*

Advances in the development of sensing devices and communication technologies have led to the development of wider and heterogeneous sensor networks (SN) in which sensor nodes are capable of monitoring, processing, storing, and transmitting physical parameters to central nodes of the networks, allowing SA systems to analyze and process this data to obtain higher-level information. Due to their nature, however, sensing devices and communication networks are inherently characterized by resource constraints: low processing power, low storage capacity, limited battery power, and limited communication bandwidth. Such limitations may degrade the capabilities and the performance of sensor networks (that is, they contribute to message loss, delays, disruptions, missed readings, etc.), and contribute to decreased quality of service and quality of information of sensor data [232, 233], with a negative impact on the applications which in turn contributes to a deterioration of the level of SA gained by the users.

Reliability of information is crucial to the process of SA assessment: missing data as well misperception of data are among the main causes of errors at Level 1 SA [14]. Consequently, improving the perceived quality of information as well as avoiding or correctly reconstructing missing sensor readings is an important task for SA systems.

In terms of the quality of the information gathered by sensors, user requirements are volatile and differ from user to user. Different kinds of applications have different quality requirements, even when using the same SN [234]. Also the level of information quality achievable depends on operating conditions such as network and environmental conditions [235]. Accordingly, SN functional operations as well as the sensor data management framework should be designed to take into account fluctuating operating conditions and changing user requirements with respect to information quality. According to Sachidananda et al. [235], we refer to quality of information (QoI) as the quality perceived by the user concerning the received information, which may fully accomplish the user's volatile requirements. QoI can be defined in terms of different QoI metrics, such as timeliness of data, accuracy, precision, latency. QoI metrics may have different importances based on the context and on non-functional requirements of the application. For instance, the timeliness of data is very important in emergency response applications, while information security is more important in health monitoring.

Ultimately, the management of sensor data in SA must take into account the requirements of different users in terms of quality metrics. The approach used for the management of the sensor lifecycle, data acquisition, and the processing and visualization of sensor readings should be aware of such heterogeneous quality requirements, and attempt to provide users with sensor data that satisfies each individual request in order to maintain high level of SA. This also requires the opportunity to deal with missing readings while simultaneously improving the perception of quality by the users to avoid errors at Level 1 SA.

In this chapter we describe an approach for quality-aware sensor data management which integrates a process for evaluating the quality of sensor data according to user requirements with a technique for estimating missing sensor data using association rules. Using this approach, each user (or application) receives only those readings that meet the requirements he/she has indicated and, when this is not possible, the user receives an estimated value. This measurement is estimated and reconstructed by means of association rules, which are mined using a variant of the approach of Hong and Wu [236]. The approach combines several intelligent techniques: i) semantic technologies for representing sensors, sensor data, and the quality thereof, thus providing a common data model which is independent of the specific data format and which enables the use of inference and reasoning processes on the gathered data; ii) computational intelligence approaches (for example, fuzzy logic) to evaluate the perceived quality with soft and relaxed constraints instead of strict thresholds as in traditional approaches; iii) an association rule mining technique to compensate for missing values.

Parts of this chapter have been previously published in:

- Giuseppe D'Aniello, Matteo Gaeta, Tzung-Pei Hong. Effective Quality-aware Sensor Data Management, *Accepted for publication in* IEEE Transactions on Emerging Topics in Computational Intelligence. ISSN 2471-285X

5.1 An Approach for Quality-aware Sensor Data Management

The overall approach consists of two main phases, as shown in Figure 5.1. In the first phase, the user sets the desired quality

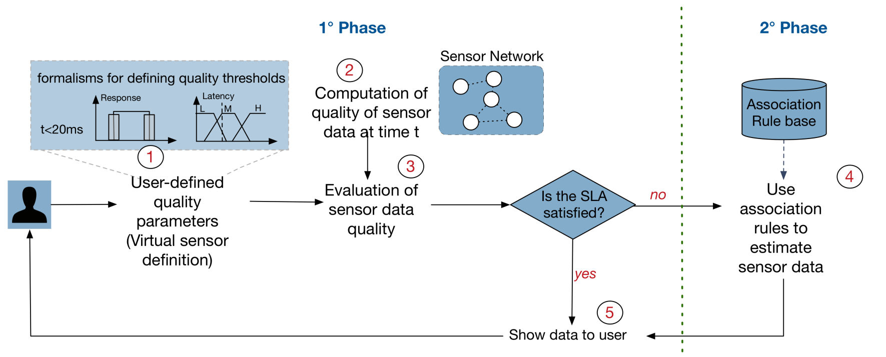


Figure 5.1 Overall approach for quality-aware sensor data management

requirements for each specific sensor or for a group of homogeneous sensors (for example, the subset of temperature sensors in the network, the subset of sensors which measure the presence of people, etc.) by defining a *virtual sensor*. A virtual sensor is a software abstraction of a real sensor. It can be seen as a sensor that provides data that satisfies the user's desired minimum quality requirements. Each virtual sensor uses the data gathered by the real sensors. By defining a virtual sensor, each user specifies his/her quality requirements for the corresponding real sensor. The virtual sensor attempts to meet the quality requirements by providing the user with:

- sensor data from the real sensor if this meets the specified quality requirements;
- an estimated value when the real sensor measurement does not meet the quality requirements;
- a reconstructed value when the real sensor is not providing measurements (for example, sensor failure, communication problems, etc.).

The advantage of using virtual sensors relies in the capability of abstracting from the physical issues and limitations of real devices. In such a way, the SA systems can use the sensors on a logical and more abstract level in which some value added services can be exploited. As an example, let us consider a physical temperature sensor and an application that needs to know the updated value of the temperature each 30 seconds. In case of communication issues, noises or sensor malfunctions, some readings can be missing. In such cases, the application should have some embedded capabilities to deal with these issues. The presence of a virtualization layer of the physical sensors overcomes such kind of issues: a virtual sensor implements all the functionalities needed in order to ensure the application with the data it needs, trying to guarantee all the quality requirements.

When the quality requirements are not satisfied by the current reading, the second phase of the approach is used (Figure 5.1). In this phase, association rules among sensor readings are exploited to compensate for missing data with suitable substitutes that have the desired level of quality.

5.2 A Virtualized Quality-aware Sensor Network

Let us consider a sensor network consisting of a set of heterogeneous physical sensors $S = \{s_1, ..., s_n\}$ as illustrated in Figure 5.2. Each sensor is characterized by several quality parameters that together represent the overall quality of the sensor itself. These parameters differ for each kind of sensor (and generally also for each specific sensor). Each physical sensor $s_i \in S$ has a set of quality parameters $Q_{s_i} = \{qp_i^1, ..., qp_i^{k_i}\}$. For instance, a temperature sensor $s_t \in S$ can have three quality parameters: $qp_t^1 =$ accuracy, $qp_t^2 =$ response time, and $qp_t^3 =$ latency.

We define a virtualization layer on the physical sensor network, namely the *virtualized quality-aware sensor network*, wherein each user defines his/her own virtual sensors on the top of the available physical sensors. A virtual sensor allows the user to define the quality parameters that he/she is interested in with respect to a physical sensor in the network and to specify also the minimum thresholds of satisfiability for each of the selected parameters. As a result, for a physical sensor network containing n sensors, we

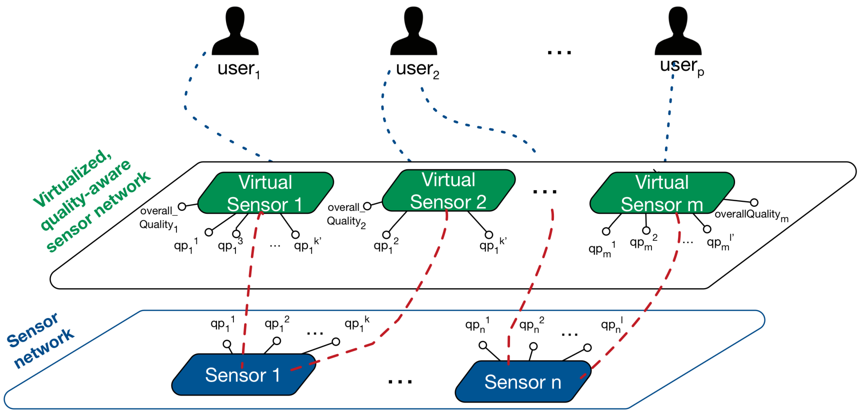


Figure 5.2 Virtualized quality-aware sensor network

define a virtualized sensor layer consisting of a set of virtual sensors $VS = \{vs_1, ..., vs_m\}$. Generally the number of virtual sensors is greater than the physical sensors ($m > n$), as the set of users $U = \{u_1, ..., u_p\}$ may define more than one virtual sensor for a given physical sensor (see also Figure 5.2).

Let us consider a virtual sensor $vs_x \in VS$ that is defined for the physical sensor $s_y \in S$. The virtual sensor has a set of quality parameters $Q_{vs_x} = qp_x^1, ..., qp_x^{k'}$ that are a subset of all the quality parameters of s_y (that is, $Q_{vs_x} \subseteq Q_{s_y}$ and $k' \leq k$).

A user $user_z \in U$ who has defined the virtual sensor vs_x indicates the thresholds of minimum satisfiability for each quality parameter $qp_x^j \in Q_{vs_x}$ with $j = 1, ..., k'$ by using a specific formalism (that is, numerical threshold, fuzzy sets, interval-valued fuzzy sets, rough sets). In Section 5.2.3 we describe an instantiation of the proposed approach using fuzzy sets to define the virtual sensors' quality parameters.

Lastly, $Overall\_quality_x$, an overall quality parameter, is associated with each virtual sensor $vs_x \in VS$, representing a global quality index for the virtual sensor with respect to the parameters considered by user $user_z$. Data gathered by the virtual sensor with an overall quality that is too low (that is, below the user's defined threshold) is discarded for the user who has defined it. Note that data gathered by the same physical sensor s_y can be discarded for a specific user $user_z$ while it can be presented to another user $user_t$, depending on their quality requirements.

5.2.1 Semantic Model of the Virtualized Quality-aware Sensor Network

The virtualized quality-aware sensor network relies on a data model that describes both physical sensors and their characteristics, the set of virtual sensors (including the quality metrics), and the characteristics of the environment in which the physical sensors are deployed. The use of semantic technologies for the data model provides us with significant benefits. Indeed, the goal of the semantic model is to establish an effective interoperability layer which eases the acquisition, collection, and processing of sensor data. In addition, it provides a common data model for representing data coming from heterogeneous sensors. Lastly, the semantic model provides a mechanism to guarantee flexibility and enable the whole system to react or adapt to possible modifications (for example, adding new virtual sensors, changing the associations between virtual and physical sensors, etc.).

The semantic model (depicted in Figure 5.3) reuses and extends the description ontology for the Internet of Things (IoT) proposed in [237]. This ontology was designed by integrating and extending existing ontologies for IoT, such as the Semantic Sensor Network Ontology (SSNO) [238], and ontologies for semantic web services like OWL-S [239]. The ontology consists of different modules, among those mentioned here: *IoT Services* for modeling services and their capabilities, *IoT Resources* and *Observation & Measurement* modules for representing sensors and their measurements, *QoS and QoI* for representing important concepts about the quality of the information provided by the sensors. Further details on the ontology can be found in [237].

For our semantic model, we are mainly interested in the *IoT Resources* module for representing physical and virtual sensors, and in the *QoS and QoI* module for representing information on

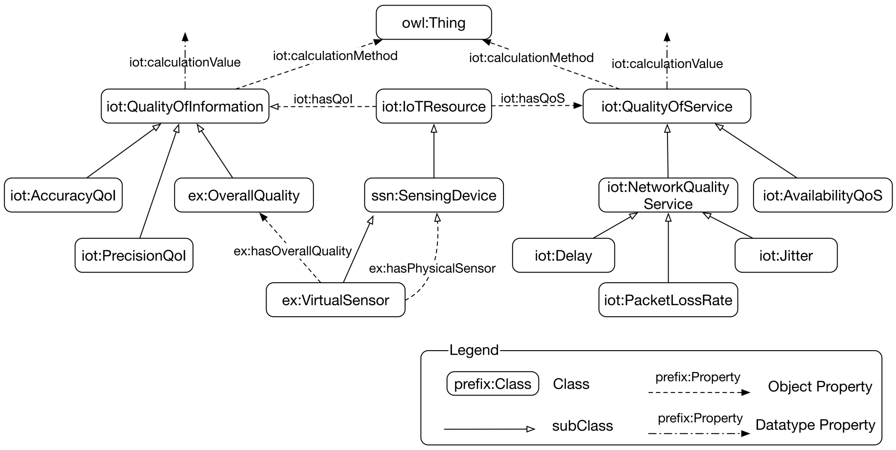


Figure 5.3 Semantic model for quality-aware sensor management

the quality of the sensor data. The main classes of this module are shown in Figure 5.3. The physical sensors and the capabilities, properties, and characteristics of their observations are represented using the main classes of SSNO. In particular, each sensor is represented as an instance of the `ssno:SensingDevice` class, which is a subclass of `iot:IoTResource`. We have extended such classes to define the class `ex:VirtualSensor` as a subclass of `ssno:SensingDevice`. Instances of this new class are used for representing the virtual sensors defined by the users. The quality parameters of the virtual sensor are specified by means of the `hasQoI` and `hasQoS` object properties. Such parameters must be a subset of the properties that have been specified for the corresponding physical sensor. The association between the virtual sensor and the physical sensor is represented by using the `ex:hasPhysicalSensor` property. For each quality parameter, the IoT Ontology foresees the definition of a function for computing its current value (represented by means of the `calculationMethod` object property) which is described by the `calculationValue` datatype property. The subclasses of the `QualityOfInformation` and `iot:QualityOfServices` classes represent quality dimensions suvh as accuracy (`iot:AccuracyQoI`), precision (`iot:Precision-QoI`), and delay (`iot:Delay`). New parameters can be considered

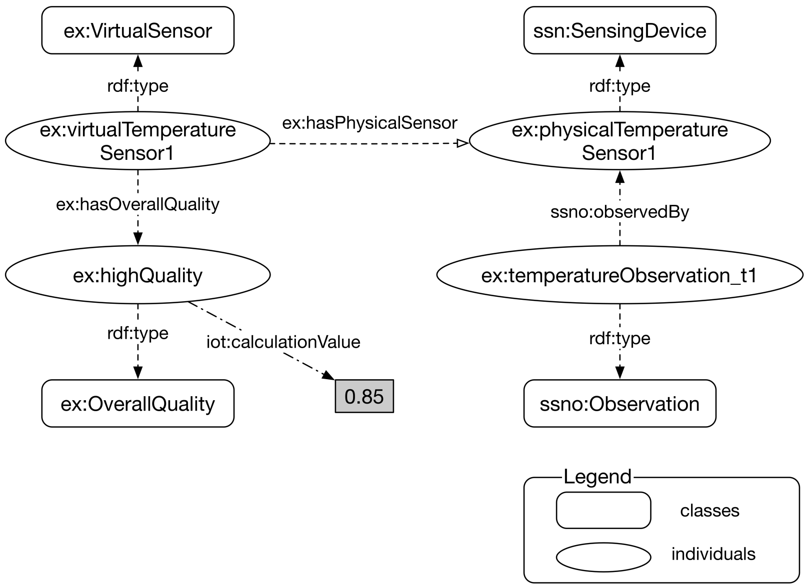


Figure 5.4 Fragment of the ontology representing the addition of a new observation and the overall quality value

by extending these classes. The class `ex:OverallQuality`, which we have defined as a subclass of `iot:QualityOfInformation`, stores the result of the quality evaluation process (Section 5.2.2).

As an example, Figure 5.4 provides a fragment of the ABOX (that is, the extensional knowledge of the knowledge base) when a new measurement is gathered by a physical temperature sensor (`ex:temperatureSensor1`). In this case, we add a new observation (`ex:temperatureObservation_t1` which is a subclass of the `ssno:Observation` class that represents the sensor measurements) and we also add the value of the overall quality of the related virtual sensor (via the `ex:hasPhysicalSensor` property). The value of the overall quality parameter is computed using the process described below.

5.2.2 Quality Evaluation of Virtual Sensors

The overall quality of a sensor reading can be evaluated by considering: i) quality parameters which can be directly measured (for example, packet loss ratio, jitter), ii) characteristics of the sensors which can be used for evaluating its quality (for example, manufacturer-declared characteristics such as accuracy, precision, max load), and iii) context information which can influence sensor quality (for example, operating temperature, lightning condition, time since last calibration).

Figure 5.5 depicts the process for evaluating the overall quality of a measurement of a single virtual sensor. The process evaluates separately the value of each quality parameter $(qp_x^1, ..., qp_x^{k'})$ when a new measurement is made available by the physical sensor. These are the main steps of the process:

1. In the first step ("*gathering data about* qp_x^k") all the information related to each quality parameter qp_x^k of the virtual sensor vs_x, including contextual data that may influence the quality, are collected from the environment. The information needed to compute values of the quality parameters are described in the semantic model.

2. In the second step of the process ("*computing* qp_x^k *value*"), the gathered data is processed according to the appropriate calculation method for computing the current value of each quality parameter qp_x^k.

3. Such values are aggregated to obtain an overall quality index. The way this index is computed depends on the specific formalism we use to define the quality requirements.

4. The overall quality value is stored in the ABOX of the semantic model and can be used by applications and services to apprise users of the current quality of the data.

5. At the same time, the overall quality is compared with the minimum level of acceptance defined by the users in the service level agreement (SLA). Again, the service level agreement can be specified by using different mathematical formalisms. When the quality is lower than the user-defined thresholds, the sensor reading is discarded.

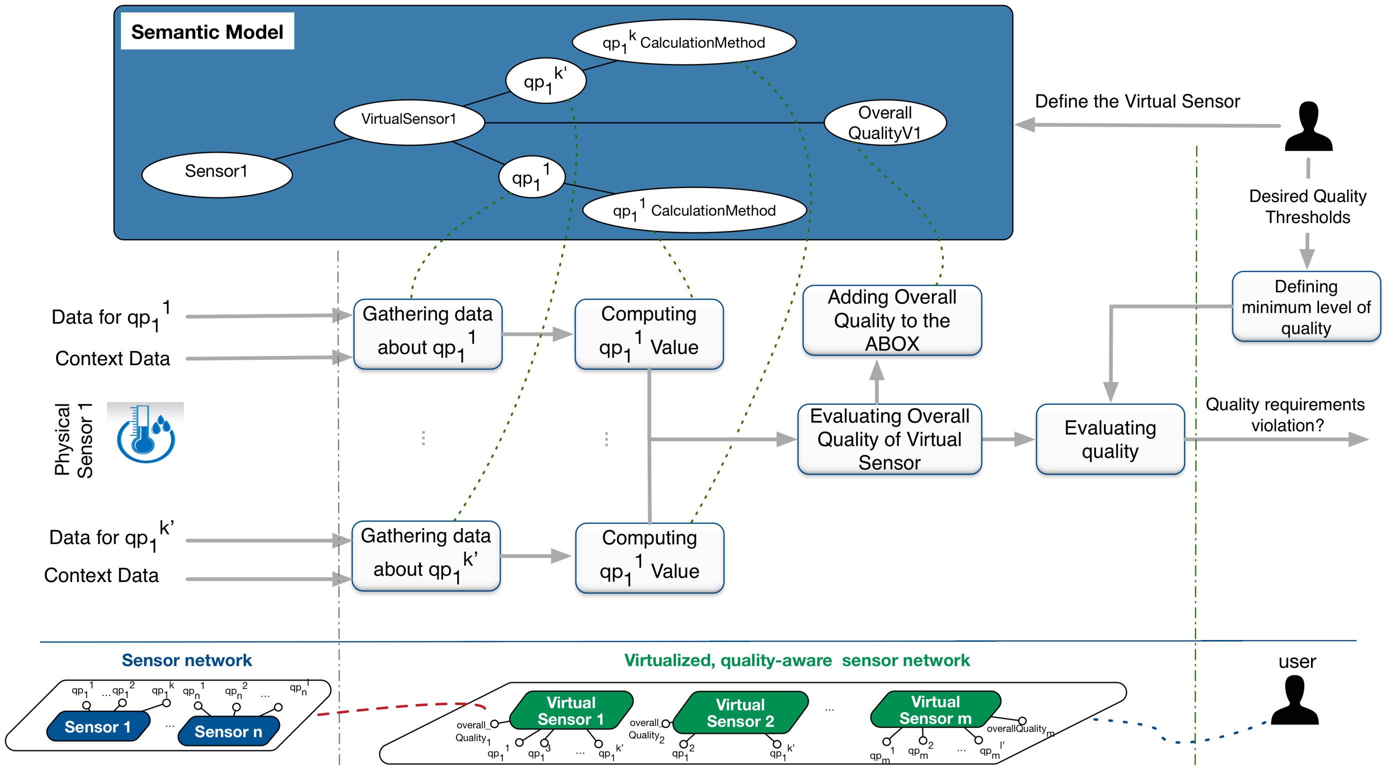


Figure 5.5 Evaluating quality of sensor data for a virtual sensor

5.2.3 Quality Evaluation using Fuzzy Sets

The process for quality evaluation supports different formalisms for the definition of the quality requirements. In the simplest case, the requirements can be defined as strict numerical thresholds on each parameter (for example, response time $\leq$ 30 ms). However, this allows no range of tolerance for each parameter, nor does it allow for the imprecise formulation and the imprecise evaluation of the quality parameters. Consequently, to permit users to define the requirements in a friendlier way and to allow for the evaluation of "relaxed" and "soft" constraints, the approach allows the users to express requirements in other ways [240], thus facilitating the partial satisfaction of user quality requirements. In particular, users may define such requirements by

- Using intervals to express a range of tolerance for each parameter (for example, response time $\leq$ 30 ms$\pm$2 ms), or even by using interval-valued fuzzy sets [241];
- Using fuzzy sets to express requirements with linguistic terms (for example, low response time and high bandwidth).

By using the latter approach, it is possible to partially satisfy users' quality requirements. We propose an illustrative example to show how the process for evaluating the quality of sensor data works when we adopt fuzzy sets to define the quality parameters of the virtual sensor. Assume that a physical temperature sensor s_t is characterized by four quality parameters:

- $q_1 = ResponseTime$: the response time, represented in milliseconds (ms);
- $q_2 = TimeSinceLastMaintenance$: as reported in [242], sensors are affected by the environment in which they are deployed. Sensors exposed to excessive heat or dust for long periods of time may degrade in performance. Thus, the time expired since the last maintenance may be considered as a parameter of quality by the user, which can not rely in a

measurement if the sensor is not maintained and calibrated for a long time;

- q_3 = *Latency*: the latency due to the network;
- q_4 = *Accuracy*: the relative accuracy, expressed as a percentage;

Let us suppose that a user defines a virtual sensor vs_t on the sensor s_t which is characterized only by two quality parameters $ResponseTime$ and $TimeSinceLastMaintenance$. Moreover, let us specify that these two parameters may change over time (they are not fixed and always equal to the value defined by the manufacturer in the data sheet). The measurements are also affected by operational conditions (for example, the sensors may operate only within a specific pressure or temperature range, or the exposure of the sensors to a polluted environment may degrade performance). Statistical or fuzzy models may be used to estimate the variation of the quality of the sensors. Such methods are formalized and represented in the semantic model via the `computationMethod` property, and allow us to evaluate, at each time instant, the current quality of the measurements.

In our approach, the user defines thresholds for two quality parameters using fuzzy sets. The user defines the two membership functions depicted in Figure 5.6 respectively for the response time and the accuracy. By analyzing such membership functions, we notice that the user considers as acceptable a value for the response time below 600 ms and as for time passed since the last maintenance a value below 60 days. Following the process described in Section 5.2.2, the two parameters are combined to evaluate the overall quality of the virtual sensors. The combination is done via a fuzzy inference system (FIS) that uses user-defined if-then rules, which implicitly express the level of the satisfiability of users' quality requirements. In particular, the overall quality is represented by the membership function of Figure 5.6c. The overall quality is normalized within the range 0 and 1. These are some rules defined by users:

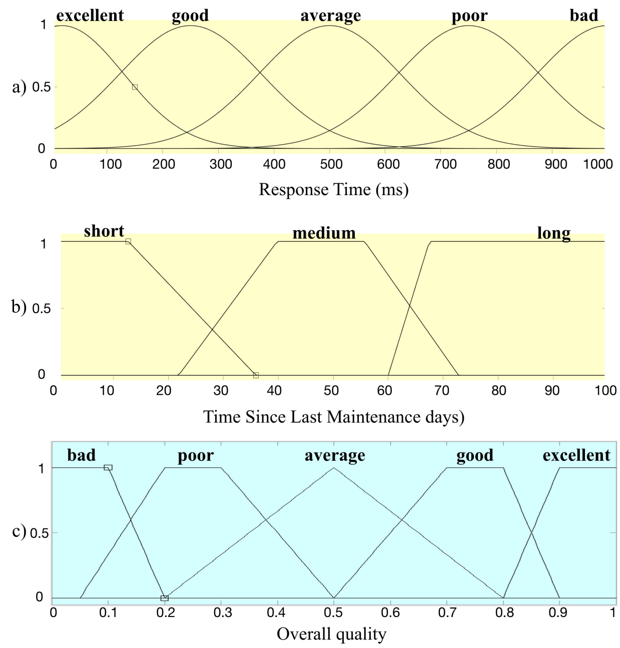


Figure 5.6 Membership functions. a) response time (input); b) time since last maintenance (input); c) overall quality (output)

1. *If (ResponseTime is excellent) AND (TimeSinceLastMaintenance is short) then (quality is excellent)*;

2. *If (ResponseTime is medium) OR (TimeSinceLastMaintenance is good) then (quality is average)*;

3. *If (ResponseTime is poor) AND (TimeSinceLastMaintenance is long) then (quality is poor).*

All the rules are applied in parallel and the fuzzy disjunction (OR) is implemented using the *max* function, while the conjunction (AND) with the *min* function. Next, the user defines a minimum level of satisfiability for the overall quality of the virtual sensor. Let us suppose that the user accepts sensor data that has an overall quality above 0.6.

Having defined the fuzzy sets and the rules, we implemented a fuzzy inference system to interpret the rules using Matlab® and

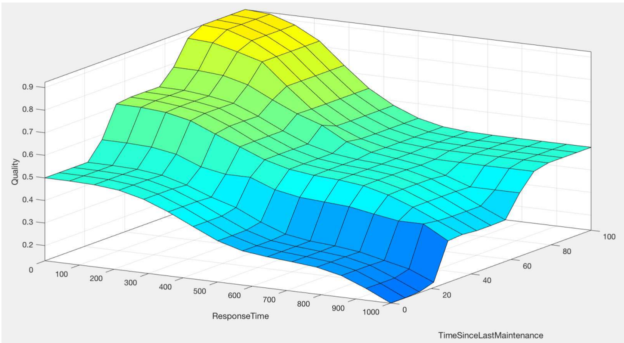


Figure 5.7 Overall quality output with respect to user-defined response time and time since last maintenance

the Fuzzy Logic Toolbox™. Figure 5.7 shows a surface map for the system and the dependency of the overall quality (output) on the response time and accuracy. In the example, we simulate and get a response time of 475 ms and that 71 days have been passed since the last maintenance. Figure 5.8 shows the rules of the implemented fuzzy inference system used in this example, together with the obtained output value for the overall quality, which in this case is 0.55. This value is below the threshold specified by the user and thus the current measurement gathered by the sensor is discarded due to its low quality (only for that user).

5.3 Data Imputation with Association Rule Mining

In order to deal with missing sensor values in the SN we use an association rule mining technique, the aim of which is to identify the association rules among the readings gathered by the sensors and to exploit them to estimate the missing value. Many different techniques for estimating missing sensor data are available [243]. The simplest approach is to delete the missing data before analyzing it [244], but this can obviously deteriorate the performance of applications that will use such data. Other techniques use statistical and machine learning approaches such as average substi-

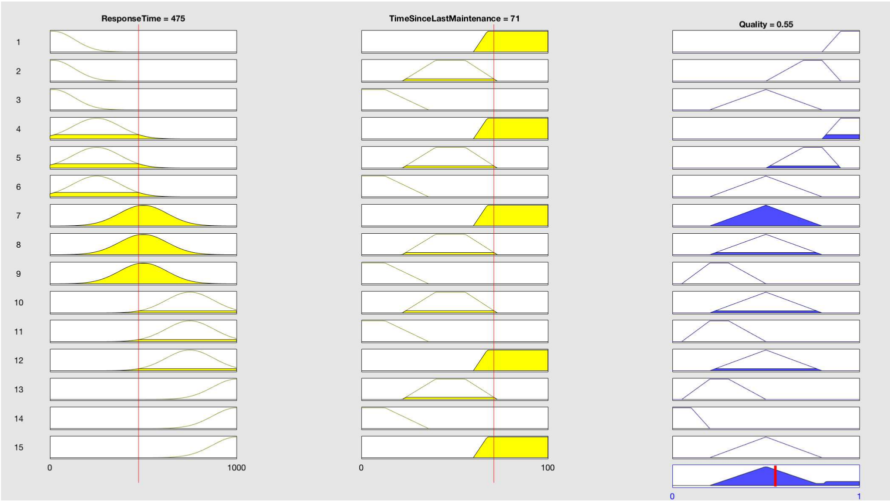


Figure 5.8 Rules of the fuzzy inference system. The input values are set as in the example

tution, imputation by regression, imputation by clustering (that is, the k-means imputation method [245]), k-nearest neighbors imputation [246], maximum-likelihood [247], and others. Other approaches exploit soft computing and computational intelligence techniques (for example, fuzzy logic [248], rough sets [249], neural networks [250], etc.) to estimate missing data or to exploit redundant sensor data to improve the quality of the measurement of environmental phenomena.

Our choice of relying on association rules is to effectively exploit the relationship among sensors in order to recover the missing values. Indeed, in case of datasets with high missing rates and with high correlation between items, the association rule mining can provide better performance when compared with traditional statistical and prediction approaches [251, 252, 236]. In the real world, most measured data always change stably; there is little mutation of environmental values between adjacent time slots. In addition, over time, environmental values are similar among some nodes. Thus, we can exploit the spatio-temporal correlation among sensor measurements to estimate missing data [253].

Let $S = \{s_1, s_2, ..., s_m\}$ be the set of physical sensors in the sensor network and assume that the time is divided into equal time slots $\{t_1, t_2, ..., t_n\}$ such that $t_{k+1} - t_k = \lambda$ for all $1 < k < n$, where λ is the size of the time slot. The interval $T_{his} = t_n - t_1$ represents the historical period of the sensor data defined during the extraction process. Moreover, let $v_i^k \in V$ be the value gathered by the sensor s_i during the time slot t_k. V is the set of values that the sensors may produce. In order to apply the association rule mining technique, V must be a finite set of elements, containing either numerical or categorical elements. When the data v_i^k is missing, we seek to identify the estimated value $\overline{v_i^k}$ thus to minimize $|v_i^k - \overline{v_i^k}|$.

The set $P = \{v_1^k, v_2^k, ..., v_l^k\} \subseteq V$ is a pattern of sensor readings, with $l \leq m$ where m is the number of sensors in S. A sensor database DS contains the sensor data and is defined as a set of epochs in which each epoch is a couple $E(E_{ts}, P)$, where P is a pattern of sensor readings that reports events within a same time slot, and E_{ts} is the epoch's time slot. An epoch $E(E_{ts}, P)$ supports a pattern $P_1 \subseteq V$ if $P_1 \subseteq P$. The frequency (or support) of P_1 in DS is defined as the number of epochs DS that support it:

$$Freq(P_1, DS) = |\{E(E_{ts}, P)|P_1 \subseteq P\}|. \tag{5.1}$$

If the frequency of the pattern P_1 is greater than a given minimum support, then P_1 is said to be frequent. Consequently, it is possible to define an association rule as $P' \implies P''$ where $P' \subset V, P'' \subset V$ and $P' \cap P'' = \emptyset$. The support of the rule $(P' \implies P'')$ represents the support of the pattern $(P' \cup P'')$ in DS, whereas the confidence of the rule is

$$Conf(P' \implies P'') = \frac{Freq(P' \cup P'', DS)}{Freq(P', DS)}. \tag{5.2}$$

An association rule is of interest if its support is greater than or equal to a threshold min_sup and its confidence is greater than or equal to a threshold min_conf. The values for these two thresholds are set empirically. In Hong and Wu [236] both min_sup and min_conf are equal to 0.5. Moreover, in the proposed approach, such values is reduced at each iteration of the algorithm in order to find more rules for completing all the missing values, as explained in what follows.

Given a database DS with an historical period T_{his} and given a minimum support and a minimum confidence, the problem of mining association rules is to generate all the rules of interest in the database. In the case of sensor networks – in particular in our approach – it is important to recall that some sensor data can be missing, resulting in an incomplete database DS (that is, some sensor values are unknown). Some approaches for mining rules simply ignore tuples which contain missing values; however, such methods may disregard important information within the data. Other methods show promising results in using association rules as an aid to completing the missing values, and have yielded acceptable prediction accuracies. These include the robust association rules (RAR) approach proposed by Ragel and Cremilleux in [254]. The limitation of this approach is that its performance may degrade when the ratio of missing values is high, as may be in our case. The algorithm for mining association rules from an incomplete dataset proposed by Hong and Wu in [236] has shown better results when dealing with datasets with high missing data rates.

We adapt this algorithm, which was proposed for generic transactional databases, to the case of the database DS containing sensor data. In this algorithm, instead of deleting tuples with missing values, robust association rules (RAR) are found by partially disabling tuples with missing attribute values. This requires redefining the concept of support and confidence to take into account the missing value. Let $Dis(P')$ be the set of disabled (missing) data with the pattern P':

$$Dis(P') = \{P | \exists A \in P', A = ?, P' \subseteq P, P \in DS\} \tag{5.3}$$

where the symbol "?" denotes a missing attribute value. A is thus a missing attribute belonging to P'.

The RAR support for a pattern P' is defined as

$$Sup_{RAR}(P') = \left\{ \frac{|Freq(P')|}{|D| - |Dis(P' \cup P'')|} \right\}. \qquad (5.4)$$

The confidence for an association rule $P' \implies P''$ based on the RAR approach is

$$Conf_{RAR}(P' \implies P'') = \left\{ \frac{|Freq(P' \cup P'')|}{Freq(P') - |Dis(P'') \cap Freq(P')|} \right\}. \qquad (5.5)$$

Using this definition for the support and confidence of RAR, the proposed algorithm implies an iterative missing-value completion method to extract the association rules.

As mentioned, the mining association rule algorithm requires that the set of values V is a finite set. Indeed, although it may be theoretically possible to apply the algorithm to real values, a finite set of values enhances the possibility to find frequent itemsets in DS. Consequently, in the case of sensor data, it is possible to use categorical values when we represent the sensor data by means of *Observations*, which can be seen as event or as linguistic representations of a sensor value. For instance, an observation can be "High temperature" instead of the numerical value 40°C, or "Presence of people in the room", and so on. When given real numerical values, we must discretize such values to obtain a reduced, finite set of elements on which to apply the mining algorithm. In this case, we introduce a precision approximation factor $pa \in \mathbb{R}$ to discretize the attribute values. Each value v_i is rounded to the nearest multiple of pa by using the formula

$$v_i^{approx} = \lfloor v_i/pa \rceil * pa, \qquad (5.6)$$

where $\lfloor x \rceil$ represents the nearest integer of x. It is evident that such an approximation plays a fundamental role in the mining process of the association rules, as it introduces errors in the estimation of missing values.

In Section 5.4 we demonstrate that for real sensor datasets and when a substantial spatial correlation among sensors exists, such approximation does not introduce significant errors when estimating missing values – rather it leads to a low estimation error.

The Hong-Wu algorithm, adapted to the problem of real numerical data as above described, consists of three main phases. The three main phases of the algorithm:

1. The association rules are mined from the incomplete original dataset and are used to roughly complete missing values.

2. It reduces the minimum support *min_sup* and the minimum confidence *min_conf* to gather more association rules from the originally incomplete dataset to complete the rest of the missing values iteratively until there are no more missing values.

3. Association rules are mined in the completed dataset. These association rules are used to correct the missing values that have been filled into predicted values until convergence. The bad influence of the missing values wrongly guessed at the beginning is reduced.

Further details on the algorithm of Hong-Wu can be found in [236]. The proposed algorithm has two main benefits:

1. It recovers the missing values of the sensor dataset with good accuracy;

2. The identified association rules among sensors can be used to estimate future missing values.

5.4 Evaluation

As anticipated in Chapter 4, we evaluate the performance of the quality-aware sensor data management approach by measuring its accuracy in data reconstruction using a real dataset. Specifically, we evaluate the performance of the approach for the reconstruction

Table 5.1 Schema of sensor readings in the Intel Lab Dataset

date	time	epoch	moteid	temperature	humidity	light	voltage
yyyy-mm-dd	hh:mm:ss.xxx	int	int	real	real	real	real

of the missing data and by comparing it with known techniques for data imputation (such as kNN and kMeans). The experimental results show that the proposed approach has a low estimation error rate, even for high missing values rates, when a significant spatio-temporal correlation exists among the sensors.

5.4.1 Data

The experiments were conducted on the Intel Lab Dataset [4], which contains the readings from 54 sensor nodes deployed in the Intel Research Berkeley Lab. The sensors were placed in the lab according to the diagram depicted in Figure 5.9. Each entry was from a Mica2Dot sensor which measures temperature, humidity, light, and battery voltage. The schema of a reading gathered by this sensor is shown in Table 5.1. The sensors took a measurement approximately once every 31 seconds. Data were collected between February 28th and April 5th, 2004, resulting in more than 2.2 million entries. In the dataset, some epochs were missing. Moreover, some values were clearly wrong (for example, room temperature over 100°C). The dataset was preprocessed to discard data like this that seemed wrong. By analyzing the dataset, we noticed that the percentage of wrong data increased as time passed. Hence we selected as t_{his} the period between 1st March to 14th March. Moreover, in order to exploit the spatial correlation among the sensors, we selected a group of sensors in the same region. In particular, we selected the sensors in the center of the laboratory. The selected region is highlighted in green in Figure 5.9. Among the 12 selected sensors, we noticed that the sensors numbered 5 and 8 were missing many epochs. Thus we used only the remaining set of 10 sensors in the selected region. We used a time slot of $\lambda = 15$ minutes and, for each sensor, we used the average value of all the readings in a time slot. In this way, after removing the epochs in which some readings were missing, we obtained 1310

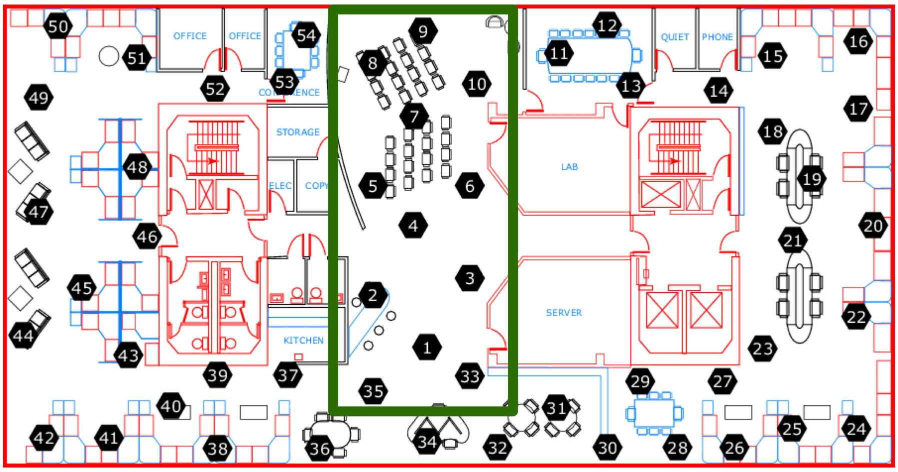


Figure 5.9 Position of sensors in the Intel Lab Dataset (from [4]).

epochs containing all the readings for the 10 selected sensors (no missing values and no missing epochs). Then we constructed two datasets for the experiments. The first contained all the temperature readings for the 1310 epochs and the 10 sensors, and the second contained both temperature and humidity readings.

5.4.2 Method

The experiments were conducted to evaluate the error in estimating the missing values in the dataset. We compared our results with four well-known techniques for data imputation. In particular, we compared it with:

- Average value imputation (AVG): missing values are estimated with the average of other attributes in the instance.
- K-nearest neighbors imputation (KNN): given the instance A which has a missing value on attribute i, it selects k other instances which have a value for the attribute i and for which the other values are most similar to A.
- K-means imputation (KMeans): first it clusters data by K-

Means with missing values, then imputes missing values with the average value of each attribute in the cluster.

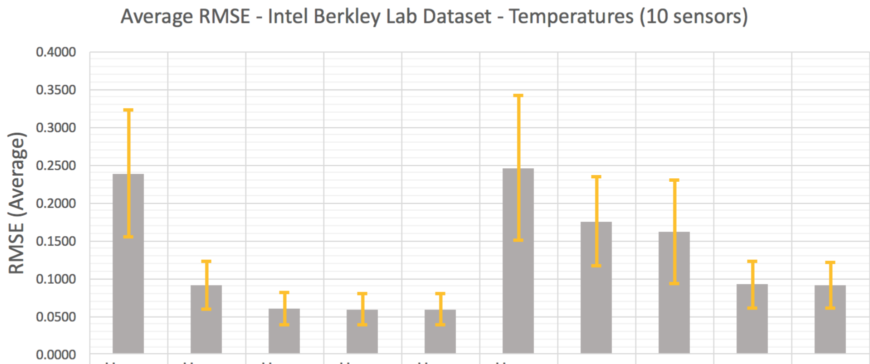
- Singular value decomposition (SVD): given a factorization of the matrix $A = U\sigma V^T$, it uses the most significant eigenvectors of V^T to linearly estimate missing values.

We implemented these imputation methods in Java by using the Statistical Machine Intelligence and Learning Engine (Smile) library [255]. The proposed association rule mining algorithm, adapted from the work of Hong and Wu [236], has been implemented in Java.

The experiments were conducted in the following way. We used the dataset (described in a previous section) as the ground truth when comparing the five techniques. To evaluate their ability to estimate the data, we simulated the presence of errors, malfunctions, and reading dropouts by randomly eliminating some readings from the dataset. Afterwards we applied the five techniques on this dataset. We compared the dataset reconstructed by each technique with the ground truth. We evaluated the estimation error as the root mean square error (RMSE), calculated as

$$RMSE = \sqrt{\frac{\sum_{k=1}^{n}\sum_{i=1}^{l}(v_i^k - \overline{v_i^k})^2}{n * l}}, \tag{5.7}$$

where v_i^k represents the value for the attribute i (that is, the sensor i), with $i = 1$ to l, at time k (that is, the epoch number k) with $k = 1$ to n in the ground truth, while $\overline{v_i^k}$ corresponds to the value estimated by the considered technique for the same attribute i at the same time k. For the dataset containing only the temperature, we have $l = 10$ and $n = 1310$; for the dataset containing both temperature and humidity readings, we have $l = 20$ and $n = 1310$. To evaluate the robustness of the techniques, we simulated missing data error rates of 5%, 10%, 20%, 30%, 40%, and 50% and compared the resulting RMSE values. Each experiment is executed 10 times for each data error rate, and the average RMSE value (and the standard deviation related to this average) is considered.



	Hong-Wu 1.0	Hong-Wu 0.5	Hong-Wu 0.25	Hong-Wu 0.10	Hong-Wu 0.05	Hong-Wu 0.01	Average	Knn	Kmeans	SVD
RMSE	0.2390	0.0912	0.0601	0.0592	0.0592	0.2465	0.1755	0.1624	0.0920	0.0917
Std. Dev.	0.0839	0.0310	0.0211	0.0206	0.0206	0.0957	0.0588	0.0685	0.0304	0.0300

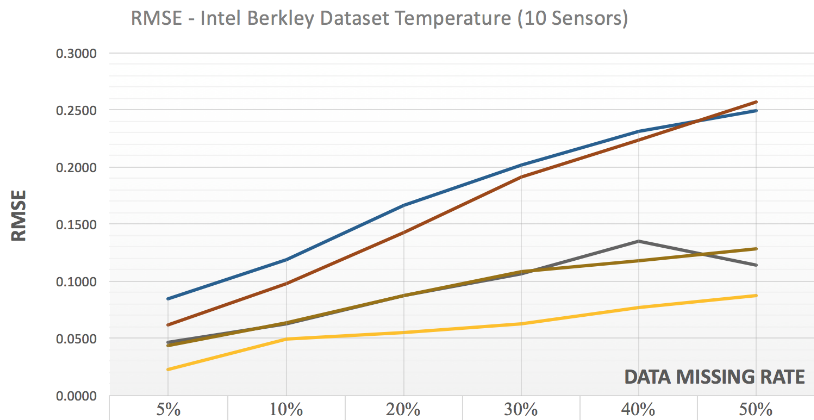
Figure 5.10 RMSE for data imputation using Hong-Wu algorithm (with various precision factors) compared with other techniques

The number of repetitions is set to 10 because we notice that, for the considered datasets and the considered techniques, the average values of RMSE converges (i.e., there are no many differences in the average values when performing more repetitions).

Lastly, to evaluate the influence of the approximation factor pa used when discretizing the numerical sensor readings, we used different precision factors (pa): $1°C$, $0.5°C$, $0.25°C$, $0.1°C$, $0.05°C$, and $0.01°C$. Even in this case, we execute 10 repetitions for each precision factor. Thus, for each of the two datasets, to identify the precision factor that best minimizes the RMSE, we executed the proposed algorithm with these six different precision factors to discretize the values using Equation (5.6). We compared the results with the other four techniques that were executed on the dataset with the original numerical values appearing in the Intel Lab Dataset.

5.4.3 Results and Discussion

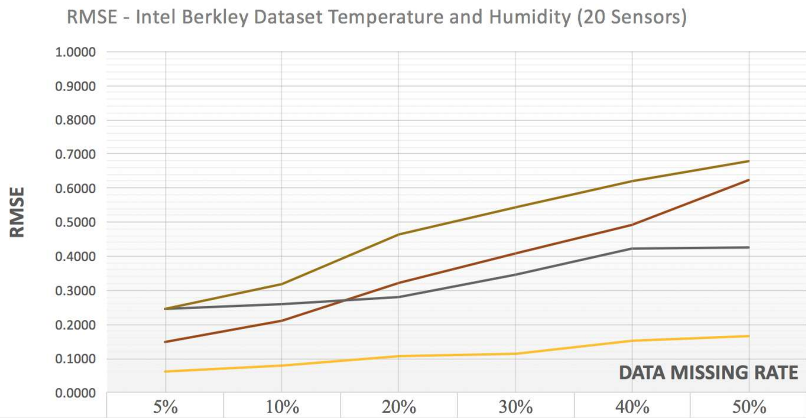
In the first experiment, we compared the RMSE of the different techniques applied to the dataset containing only the tem-



	5%	10%	20%	30%	40%	50%
Hong-Wu 0.10	0.0229	0.0492	0.0556	0.0631	0.0769	0.0873
Average	0.0847	0.1192	0.1665	0.2020	0.2311	0.2495
Knn	0.0617	0.0983	0.1424	0.1916	0.2237	0.2569
Kmeans	0.0462	0.0623	0.0878	0.1067	0.1348	0.1140
SVD	0.0440	0.0636	0.0879	0.1080	0.1181	0.1285

Figure 5.11 RMSE for different missing data error rates for the evaluated techniques on the temperature dataset.

perature readings (10 sensors). We simulated different missing data rates to evaluate the robustness of each technique. Figure 5.10 shows the average RMSE and standard deviation for the various missing data rates for each technique. Notice that the proposed algorithm (termed "Hong-Wu" in the figure) was applied with the above-mentioned precision factors. In Figure 5.10 "Hong-Wu $<n.nn>$" refers to the Hong-Wu algorithm applied to the dataset with an approximation precision factor of $<n.nn> \in [1.0, 0.5, 0.25, 0.10, 0.05, 0.01]$. Note that with a precision factor $0.5 \leq pa \leq 0.05$, the Hong-Wu algorithm yields an RMSE lower than the other techniques. In particular, the lowest RMSE is both for $pa = 0.10$ and $pa = 0.05$. Figure 5.11 shows the RMSE metrics for each technique at the different missing data rates. In this figure, we show only the results of the Hong-Wu Algorithm with $pa = 0.1$. The proposed algorithm outperforms the other techniques for all the missing data rates. We repeat the same process for the second dataset, which contains both the temperature read-



	5%	10%	20%	30%	40%	50%
Hong-Wu 0.20	0.0605	0.0800	0.1080	0.1147	0.1533	0.1667
Average	1.9259	2.6967	3.7668	4.5014	5.1195	5.5972
Knn	0.1474	0.2113	0.3213	0.4081	0.4932	0.6232
Kmeans	0.2441	0.2594	0.2798	0.3445	0.4225	0.4263
SVD	0.2445	0.3187	0.4647	0.5444	0.6203	0.6802

Figure 5.12 RMSE at different missing data error rates for the temperature+humidity dataset (20 sensors)

ings and the humidity readings from the 10 selected sensors. In this case, the best performance of the Hong-Wu algorithm is at $pa = 0.2$. Figure 5.12 shows the RMSE for the different missing data rates of all the techniques. Again, the Hong-Wu Algorithm outperforms the other techniques at all the different missing data rates. The obtained results show that the Hong-Wu Algorithm exploits spatio-temporal correlations among sensor readings to estimate missing values with good performance. It is also robust with respect high missing data rates in the dataset, showing low RMSE values. One drawback when dealing with real values is how to determine the precision factor for discretizing the sensor readings to obtain a finite set of attribute values, since this influences the final RMSE. In this chapter, we have identified this factor with different tests and experimental attempts. When the sensor data are categorical (for example, binary values, linguistic values, etc.), this is not an issue, so the technique can be directly applied to the available data, simplifying the estimation process.

5.5 Summary

As we discussed in Chapter 2, level 1 SA errors are responsible for more than 75% of errors in SA. Missing data and reliability of information represent serious issues for level 1 SA. Thus, addressing such issues is a must for improving the situation awareness. In this chapter, we have proposed an approach for the virtualization of sensor network which allows to manage sensor data with a focus on the quality of information. By means of this approach, users define virtual sensors to indicate which kind of data they wish to receive from the network and which are the desired quality requirements. The virtualization of the sensor network is realized by leveraging on a semantic model that allows the upper-level applications to use a common, formal, shared, interoperable and domain-independent model of data for processing data gathered by physical sensors, thus avoiding traditional low-level issues. Moreover, the virtual sensor provides the applications with another valued added service that is the capability of estimating a missing sensor reading. Such capability is implemented by using an association rule mining algorithm which exhibits good performance even in case of many missing data in the training set. Specifically, in the reported experiments, the association rule mining outperforms other data imputation techniques in cases where a high spatio-temporal correlation exists between sensor readings.

Chapter 6

Adaptive Goal-driven Situation Management

"Concentrate all your thoughts upon the work at hand. The sun's rays do not burn until brought to a focus."

— Alexander Graham Bell

According to the IEEE SMC Technical Committee (TC) on Cognitive Situation Management[1], the term *Situation Management* identifies collectively all the operations for supporting situation awareness, prediction, reasoning and control. A similar definition of situation management has also been proposed by Jacobson, Buford and Lewis in [256]. The focus is on the management viewpoint, according to which it is important to understand the situations involving interdependent dynamic entities with complex relations, to recognize emerging trends and potential threats, and to undertake required actions. Continuing with the definition provided by the TC, the understanding of dynamic situation requires complex cognitive modeling of situations, building formal situation models, and continuous sensing, perception, and comprehension of signal and human intelligence events and reports, and integrating

[1]http://www.ieeesmc.org/technical-activities/cybernetics/cognitive-situation-management

this data into suitable presentations for supporting human and computational understanding of situations.

In this chapter, we propose an adaptive approach, namely the Adaptive Goal-driven Situation Management approach (AGSM), based on a formal and semantic modeling of goals and situations, which contribute to the situation management with a specific focus on addressing these SA demons: attentional tunneling, data overload, out-of-the-loop, misplaced salience. Through an ontological model of users' goals and situations, AGSM concretely implements some of the principles envisioned by Endsley in order to address such demons. Specifically, it supports the users in the alternation between the goal-driven and data-driven information processing, in order to avoid losing the global view of the environment while focusing on what really matters according to their goal. Furthermore, it supports the users in considering the goal which is the most important due to the surrounding conditions. In such a way, the approach contributes to soften the SA errors caused by the aforementioned demons, like data overload, attentional narrowing, wrong or poor mental models, difficulty in maintaining multiple goals.

In what follows, first we analyze the design principles and best practices usually exploited to address the attentional tunneling, data overload and misplaced salience demons. By leveraging on such principles and best practices, we define the AGSM approach by describing: i) the computational approach supporting the alternation between goal-driven and data-driven information processing; ii) the semantic model of goals and situations which sustains the AGSM; iii) the goal selection approach based on goal desirability to support users in focusing on what really matters at a given time; iv) the reinforcement learning technique to adapt the goal selection approach to the users' preferences.

Lastly, as anticipated in Chapter 4, we evaluate the AGSM approach by using SAGAT to measure the improvement in the SA and by executing a numerical simulation to measure the improvement in the decision making performance.

Parts of this chapter have been previously published in:

- Giuseppe D'Aniello, Vincenzo Loia, Francesco Orciuoli. A multi-agent fuzzy consensus model in a Situation Awareness framework, In Applied Soft Computing, Volume 30, 2015, Pages 430-440, ISSN 1568-4946.
- Giuseppe D'Aniello, Angelo Gaeta, Matteo Gaeta, Mario Lepore, Francesco Orciuoli, Orlando Troisi. A new DSS based on situation awareness for smart commerce environments (2016) Journal of Ambient Intelligence and Humanized Computing, 7 (1), pp. 47-61.
- Giuseppe D'Aniello, Angelo Gaeta, Vincenzo Loia, Francesco Orciuoli. Integrating GSO and SAW ontologies to enable Situation Awareness in Green Fleet Management. 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, CogSIMA 2016, art. no. 7497801, pp. 138-144.
- Giuseppe D'Aniello, Vincenzo Loia, Francesco Orciuoli. Adaptive Goal Selection for improving Situation Awareness: the Fleet Management case study, In Procedia Computer Science, Volume 109, 2017, Special Issue 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16-19 May 2017, Madeira, Portugal, Pages 529-536, ISSN 1877-0509.
- Gianpio Benincasa, Giuseppe D'Aniello, Carmen De Maio, Vincenzo Loia, Francesco Orciuoli. Towards perception-oriented situation awareness systems (2014) Advances in Intelligent Systems and Computing, 322, pp. 813-824.

6.1 Motivations

In the context of operational SA, the way by which information is presented to the human operators via the interface influences SA [15]. It is straightforward that the presentation is actually just the last step of a long chain of information processing phases, ranging from gathering data from sensors to the identification of the situations. The proper design of information processing approaches and SA systems in this context is crucial to correctly present information to users and reducing the issues related to SA demons.

Endsley and Jones in [15] proposed a set of design principles and recommendations that play a major role for addressing attentional tunneling, data overload and misplaced salience demons. We briefly reports those on which we founded the AGSM approach.

A first general principle suggests organizing the information in terms of the operator's major goals, instead of presenting them based on the sources of information (as usually is done in traditional decision support systems). In such a way, all necessary information to make a decision can be easily identified by the operators, reducing their overall workload. A technique for the identification of both the user's major goals as well as the SA information requirements related to them is the Goal-Directed Task Analysis (GDTA) [15], a form of cognitive analysis often used in the initial steps of designing an SA system. Another important design principle is the support to the global SA (i.e., a global picture of what is happening in the environment), which only apparently is in contrast with the previous principle that indeed suggests to organize and filter the information according to a single, specific goal. Indeed, this principle stresses the importance that the operator should always be able to have the big picture of what is happening in the environment. If the interface is designed to direct the attention only on a subset of information, the so-called attentional narrowing error may arise. The global picture and the detailed information related to the current goal of the operator should be traded off. Usually, the interface visualizes salient information related to global SA, as this kind of information is critical for accurately determining which goals should have the highest priority. Specifically, information coming from the global picture may lead the operator to reprioritize his/her goals and to switch the attention to a more important goal. Strictly related to the global SA, another design principle [15] suggests to made salient the critical cues capable of activating schemata and mental models. Indeed, such cues can support the operator in activating the right mental model, thus focusing the attention on the critical information in a given moment and allowing for reprioritizing the goals. Lastly, it is crucial to consider the alternation between goal-driven and data-driven information processing approaches. Basically, this means to combine all the aforementioned principles wisely. Organizing the information around operator goals support the goal-driven processing; the big picture for global SA support the data-driven processing by recalling the user attention on more priority goals, and the switching between goals is also supported by defining and making the critical cues salient. Thus, the key is to ensure the alternation between these two approaches to improve and maintain the SA of the operators. Following such considerations, in the AGSM approach, we concretely support the trade-off between goal-driven and data-driven with an adaptive mechanism of goal selection that suggests to the human operators which is the information that recalls his/her attention. Such mechanism allows the operators to reprioritize their goals, without losing the global picture while focusing on a specific goal, since the approach continuously and autonomously evaluates the importance of critical events and important cues of the environment to understand if the human operators should focus on a different goal.

6.2 Adaptive Approach for Goal-driven Situation Management

Before describing the AGSM approach, we propose some methodological considerations related to the design phases of SA systems willing to support the trade-off between goal-driven and data-driven information processing. Figure 6.1 depicts the main elements to take into account when defining an SA system which supports the combination of the two aforementioned processing approaches. Two aspects characterize the proposed approach:

- it combines data oriented (bottom-up) and goal oriented (top-down) analysis in order to mimic two fundamental mechanisms of human cognition: integration of parts into whole and decomposition of the whole into parts;

- it supports situation identification and adaptation to multiple views and perspectives of different human operators.

In the first phase (mostly performed by humans), the domain and the tasks, the problems and the decisions of the users, should be analyzed according to the two approaches for the analysis: top-down and bottom-up. In such a way, all the SA needs and information requirements will be identified, since by proceeding in a bottom-up phase, one will consider all the details and elements that influences SA and decisions, while when proceeding in a top-down way, one will focus on the goals (e.g., by using the GDTA technique) and will filter only the relevant information, thus avoiding to be overwhelmed by useless details. As a result of this phase, the real world problems are structured in terms of hierarchies of concepts, levels of abstraction of data and processes, set of tasks to be performed, information requirements and SA needs, as well as other information that can be processed in the next phase.

The second phase of the proposed approach (mostly performed by computational entities like software agents) relates to the combination of goal-driven and data-driven information processing to support operational SA at run-time. This phase combines situation identification, to recognize current situation with respect to goals and specific events that happen, with adaptation mechanism to support goal-changing. Notice that the results provided by the first phase cannot be directly computed by the software agents, as those results consist of human-understandable data structures.

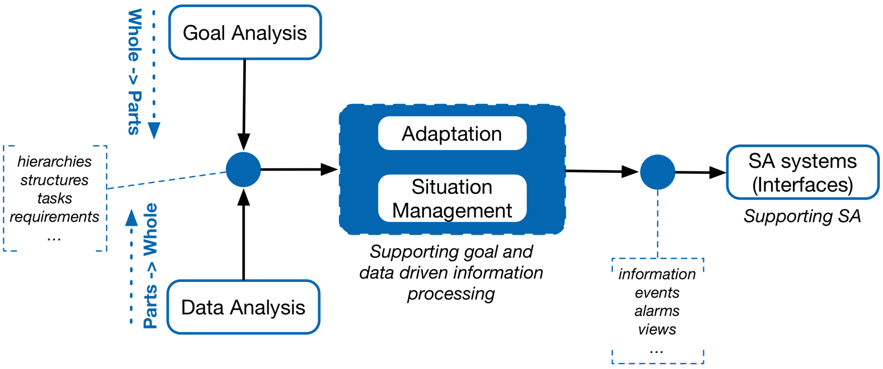


Figure 6.1 Methodological approach for the definition of SA systems supporting the trade-off between goal-driven and data-driven information processing.

Therefore, it is needed a mechanism to computationally model such data structures. The AGSM approach adopts computational ontologies to accomplish this aim. Ontologies mainly provide both an interoperable layer and shared knowledge among humans and software agents, allowing them to concretely cooperate for the realization and running of systems supporting SA effectively.

The third and last phase deals with the support to human SA, that means to present and organize the information around user goals via properly designed interface capable to draw human attention to relevant cues.

6.2.1 Functional View of the AGSM Approach

The Adaptive Goal-driven Situation Management (AGSM) approach enables the combination of goal and data driven information processing as shown in Figure 6.2, which makes evident the way by which users process information when interacting with an SA system, together with the main reasons that cause a transition from goal-driven to data-driven and viceversa. The AGSM approach exploits such transition in order to stimulate and sustain users in achieving a proper trade-off between the two modalities, without exceeding in both senses.

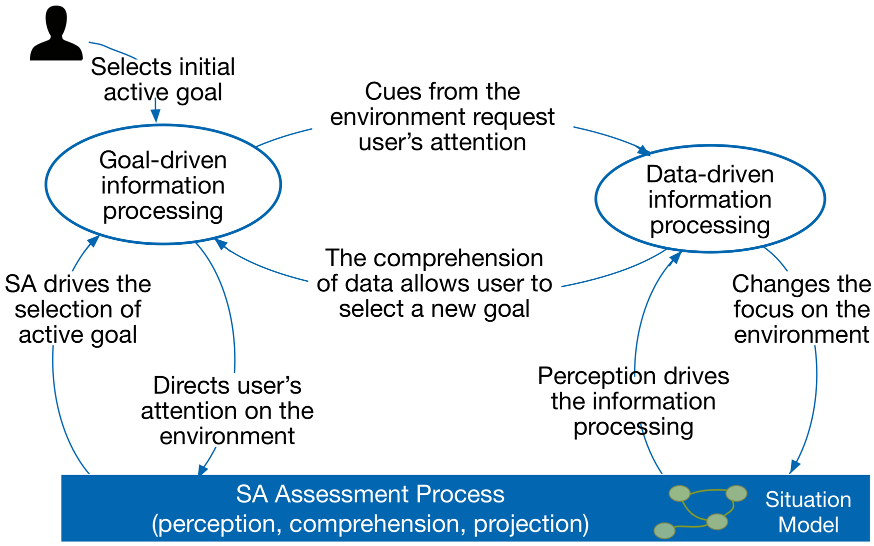


Figure 6.2 Goal-driven and data-driven information processing

Generally, users perform an initial task having a specific goal to achieve (goal-driven), and they continue to process information by considering each time a different goal. We define as the *active goal* the one the user is considering when processing information in this way. The active goal determines which elements of the environment to pay attention and influences Level 2 SA, because it has impact on the way by which users perceive and interpret information and understand current situations (the mental model of the users depends on the active goal). At the same time, being aware of the current situation helps users to determine which goal(s) should be considered afterwards for processing information. During goal-driven information processing, some elements of the environment not related to the goal may capture user's attention. For instance, alarms, flashing icons on the screen, disruptions or malfunctions, represent cues that may lead people to change their focus. When this happen, the user starts processing information from the bottom, without considering the current goals, in order to understand what is happening. This process drives the SA formation: when the users have identified what is happened, they have the ability to perform the tasks that are needed for dealing

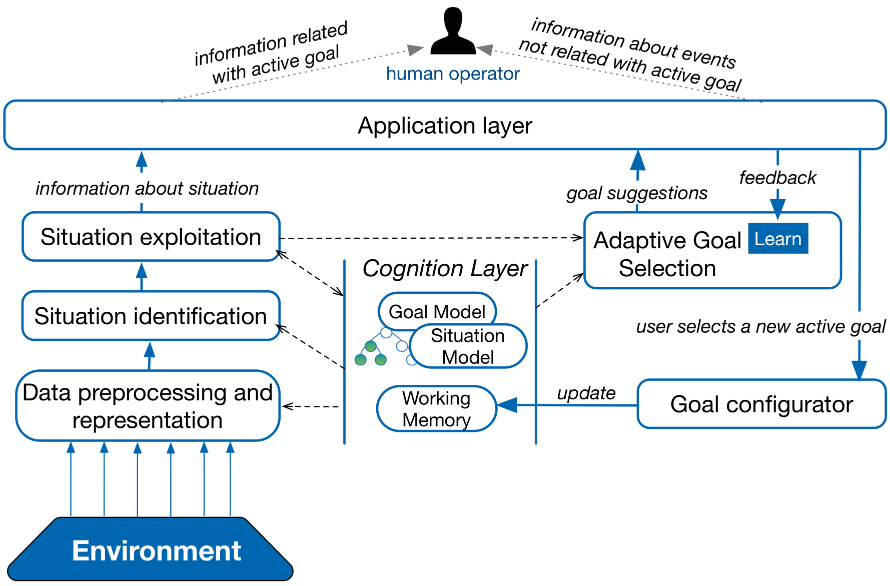


Figure 6.3 Adaptive Goal-driven Situation Management: a functional view

with the situation. This allows the users to re-prioritize the goals so to switch again to the goal-driven information processing (a more efficient way to process information).

In order to enable and sustain this alternation between goal-driven and data-driven information processing, different capabilities and functionalities are needed. In Figure 6.3 we sketch such capabilities by means of a functional view. The figure depicts the main phases and the data flows (starting from sensor data to reach human operators/users and from them back to the environment) that are involved in the process of alternation between goal-driven and data-driven.

At the bottom of Figure 6.3, data gathered from the environment is preprocessed in order to deal with uncertainty, noises, outliers. The approach for quality-aware sensor data management proposed in Chapter 5 can be exploited in this stage. The data is interpreted in order to identify the current situation (*Situation Identification*) by means of data fusion techniques. Data is classified and elaborated (according to the active goal of the user) in order to provide a representation of the occurring situation. All the possible situations are modeled in the *Cognition layer*. The Cognition layer contains the data structures and the shared knowledge useful to perform all the phases of the AGSM approach. Among the others, it contains the ontological models for representing situations, goals and sensor data. While it can seem obvious to represent situations in a computational and formal way (as also other approaches do), the AGSM approach also proposes a computational model of the user goals. This aspect is peculiarly important in AGSM since the active goal must influence the behaviors of all the functionalities of the approach to make it concretely adaptive with respect to the user needs. Based on the active goal, the information that is important for the users may change; moreover, only a subset of situations may be relevant to a specific goal; also, the way by which data is processed may change according to the goal of the user; lastly, also the decisions and the tasks the user should perform changes according to the active goal, and so should change the interface.

The Cognition layer, moreover, contains a data structure called *Working Memory* (WM) which mimics the working memory of humans since it contains the information needed in the short time to support SA. In details, WM contains the information related with the active goal, the current situation, and the necessary data to complete the user's tasks. Details about the ontological model of goals and situations are reported in Section 6.2.3.

The identified situation is further elaborated by means of reasoning and inference processes (indicated as *Situation Exploitation* in Figure 6.3), and it supports the applications and interfaces (contained in the *Application Layer*) in the exploitation and visualization of the information, concerning with the identified situation, in order to sustain users in making decisions or taking actions, thus supporting the human operators to focus only on what matters for the goal (goal-driven information processing).

At the same time, the approach will continuously monitor the environment (whatever is the goal) in order to promptly inform the users about important asynchronous events that require their attention (data-driven information processing). Specifically, the *Adaptive Goal Selection* phase depicted in Figure 6.3 considers the current situation, the active goal of the user, the data about asynchronous events, alarm conditions, in order to decide if the user should focus on a different goal. If this is the case, it suggests to the user to change the active goal by means of notifications and events (i.e., using salience of information) convoyed via the interfaces of the application layer. When the user decides to change the active goal (due to his/her willingness or because stimulated by the adaptive goal selection approach), the *Goal Configurator* updates the content of the WM with the information related with the new goal. In this way, all the other phases (situation identification and exploitation, adaptive goal selection and also the application layer) are implicitly informed of this change and can adapt their behavior accordingly. Details about the adaptive goal selection process is reported in Section 6.3. The change of the active goal performed by the user after a suggestion by the adaptive goal selection process is useful for learning if the user likes such suggestions and to adapt the future behavior accordingly. Details about the process of learning from users' feedback are provided in Section 6.3.4.

6.2.2 Goals

The set of user goals that are contained in the Cognition Layer, should be identified by analyzing the SA information needs of the operators. As described in previous section, this can be systematically achieved by exploiting the Goal-Directed Task Analysis (GDTA) [15]. GDTA focuses on the goals the operator must accomplish in order to successfully perform a task, the decisions he/she must make to achieve the goals, and the information that is needed in order to make the appropriate decisions. An example of the goal hierarchical structure obtained by following the GDTA approach is sketched in Figure 6.4, with reference to the logistic domain. Specifically, this example is related to the management of a fleet of vehicles responsible for the delivery of products, with

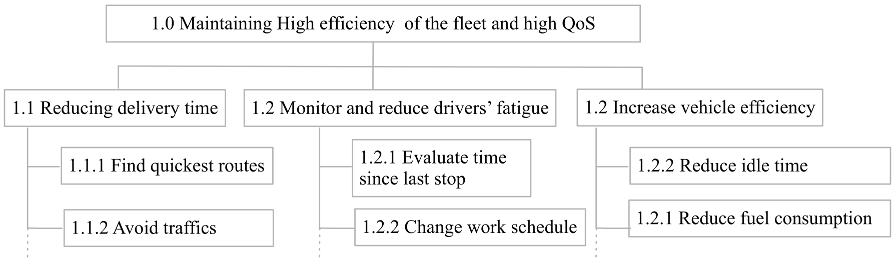


Figure 6.4 Example of hierarchy of goals identified by means of GDTA technique in the logistic domain.

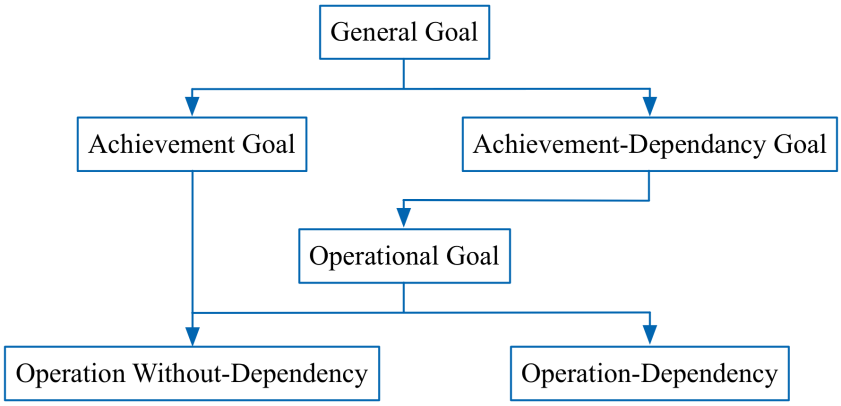


Figure 6.5 Goal classification (from [5])

the aim of increasing their efficiency. One of the result of the GDTA is a hierarchical structure of goals which is not enough formal to be processable by computational entities. Consequently, to allow software agents to interpret goals, a formal model of goals is needed. To define such a model, we need to identify all the possible type of goals to represent. To this aim, we adopt the goal classification proposed in [5] related to a goal-driven requirement analysis approach. Figure 6.5 depicts the proposed classification, which foresees the following types of goals:

- General Goals: high-level goals used to express the overall objective of the users from their point of view. Usually, these goals represent the roots of the GDTA schema.

- Achievement Goals: are intermediate goals (subgoals) which allow to achieve a general goal. Usually, they correspond to subgoals of the General goals.

- Achievement-Dependency Goal: is an Achievement Goal refined in, at least, one Operation-Dependency Goal. Thus, in order to fulfill this goal, first we need to solve the dependency among the Operational Goals in which it is refined.

- Operational Goals: represent the objectives of operations/-tasks that can be directly fulfilled by computational entities and humans.

  - Operation-Dependency: the achievement of this king goal depends on the competition of another operation.
  - Operation Without-Dependency: this goal can be directly completed, without no dependency with the achievement of other goals.

The arrows in Figure 6.5 represent the ways in which a goal can be refined. A General Goal can be refined in other General Goals or in Achievement Goals (with or without dependency), while Achievement Goals can be refined in other Achievement Goals or Operational Goals. The Operational Goal represents the lowest level of refinement, as such kind of goals are directly realized by software agents. In order to computationally represent the identified goals, a set of existing ontologies have been exploited and extended in order to build an ontological model representing goals, situations, domain knowledge and sensor data, described in next section.

6.2.3 Semantic Model of Goals and Situations

The semantic model of goals and situations contained in the cognition layer is defined by integrating two ontologies: the Goal Service Ontology (GSO) [257] and the Situation Awareness Core Ontology (SAW) [258].

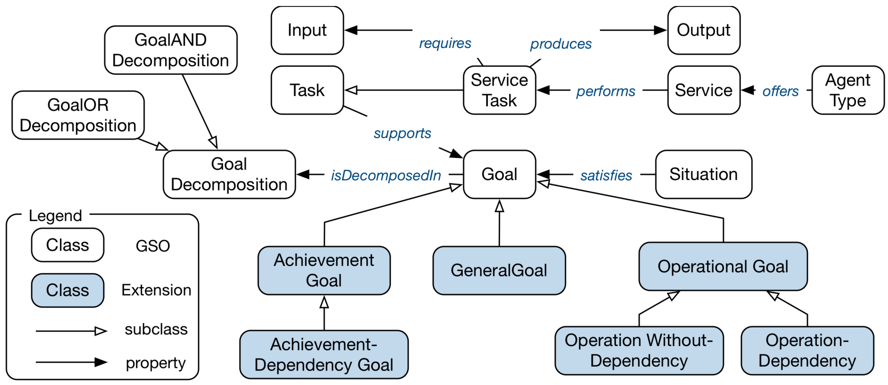


Figure 6.6 Sketch of the main classes of GSO, extended with the classification of goals.

The Goal Service Ontology (GSO) is adopted to formally model goals. A sketch of the main classes of this ontology is shown in Figure 6.6. The class `gso:Goal` represents a goal. We have extended this class with the category of goals identified in the classification schema [5]. In the figure, the hollow arrowhead represents a subsumption relation. The class `gso:Task` represents the task to execute in order to fulfill the goal it supports. Such tasks are implemented by software agents or any other computational entities capable of implementing the required services. Without loss of generality, the model represents such computational entities with the class `gso:AgentType`, which offer services (class `gso:Service`) able to perform the related task. Thanks to this model, once a new goal is active, it is possible to automatically select the tasks needed to satisfy it, and to check if such goal is realizable or other preconditions or dependencies must be satisfied first. Notice also that a Goal can be further decomposed in other subgoals (via the `gso:GoalDecomposition` class). This allows the software agents to decompose the overall (Operational-) Goal in simpler subgoals which they are able to satisfy.

Goals can be related to situations modeled by means of SAW ontology [258]. By using the object property `gso:satisfies`, it is possible to link the class `gso:Situation` with the class `gso:Goal`.

The main classes of the situation model are depicted in Fig-

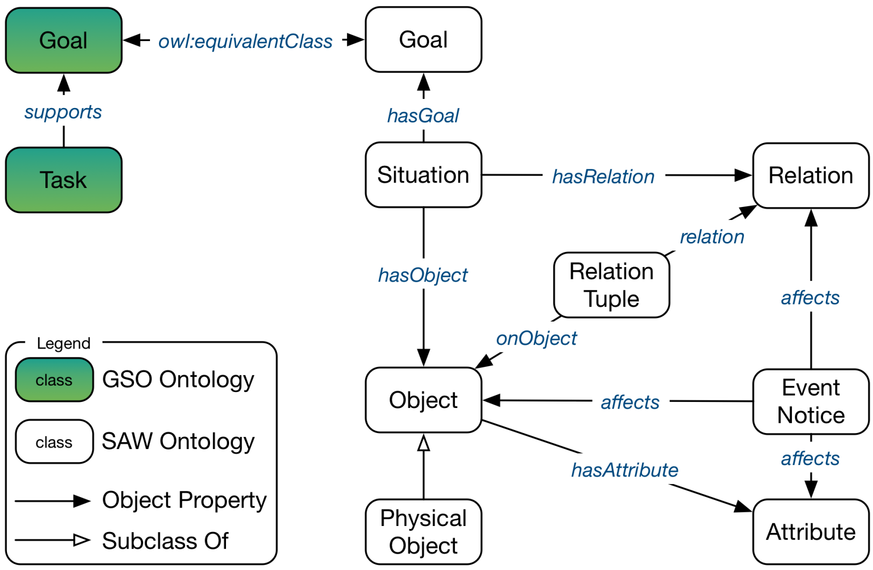


Figure 6.7 Situation Model

ure 6.7. Notice that the classes `gso:Goal` and `saw:Goal` represent the same concept (i.e., the user goal) and contain the same individuals. As a consequence, the two classes are considered equivalent (thus we have set the `owl:EquivalentClass` property between the two classes). The class `gso:Goal` allows to select the tasks (represented by the class `gso:Task`) the agents need to accomplish in order to fulfill such goal. On the other side, the class `saw:Goal` allows to select the situations (modeled by the `saw:Situation` class) that are relevant for the current user's goal. The main entities that are relevant for each situation are represented by the class `saw:SituationObjects` that can have characteristics represented by the class `saw:Attribute`. Such entities may participate in relationships represented by the class `saw:Relation`. The class `saw:RelationTuple` represents which `saw:SituationObject` participates in a `saw:Relation` and its truth value. Both `saw:Relation` and `saw:Attribute` are associated with values that can change over time, represented by the class `saw:PropertyValue`. Moreover, it is possible to associate

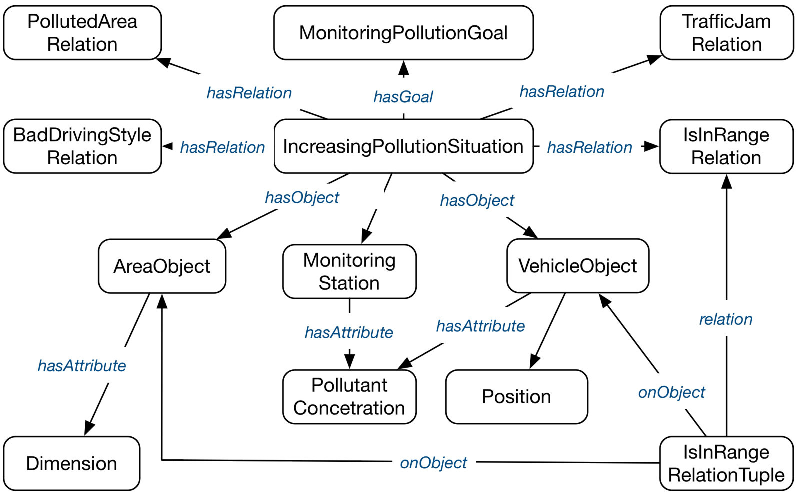


Figure 6.8 An example of instantiation of the Situation Model

to each relation a `saw:Rule`, that fires at particular events (e.g., changing of a value for an attribute, external events). The firing of the rule defines the truth value for the `saw:Relation`. Each situation has some relevant relations, indicated by the property `saw:RelevantRelations`. The verification of all the relations relevant for a situation allows verifying if the `saw:Situation` occurs in the environment, hence to identify the situation. When a specific situation has been identified, it is possible to add an instance of it to the extensional knowledge (ABox) of the knowledge base, contained in the Working Memory of figure 6.3. In this way, the software agents know at any time instant what are the occurring situations and they can act coherently with it.

As a proof of concept of the proposed model of goals and situations, Figure 6.8 shows an example of the model instantiated in the logistic domain with respect to the example of GDTS depicted in Figure 6.4, and specifically for monitoring vehicles and their pollutant emissions. Each box in the figure represents a subclass of the proposed ontological model of goals and situations. In this example, the active goal of the user is to monitor the pollution in a specific area where vehicles are in transit (`MonitoringPollutionGoal` subclass of `saw:Goal`). One of the possible situations relevant for this goal is `IncreasingPollutionSituation` (subclass of `saw:Situation`), which represents the increase of pollution in that area. In order to verify if this situation holds, three relations need to be evaluated: i) the level of pollution in the area (subclass `PollutedAreaRelation`); ii) if in the considered area there is a traffic jam (subclass `TrafficJamRelation`); iii) the bad driving style of the drivers transiting in the area (subclass `BadDrivingStyleRelation`). The relevant `SituationObjects` of the `IncreasingPollutionSituation` are the `Area` (whose main attribute is the `Dimension`), the `MonitoringStation` (whose main attribute is the `PollutantConcentration`) and the `Vehicle` (whose main attributes are the `Position` and the `PollutantConcentration`). Regarding the relation `BadDrivingStyleRelation`, in order to identify only that drivers transiting in the considered area, the subclass `IsInRangeRelation` is used. Such `Relation` is defined on two `SituationObject` (`VehicleObject` and `AreaObject`), via the `IsInRangeRelationTuple` class and `onObject` property (see Fig. 6.8). The instances of this `Relation` represent the vehicles that are in the area. Such instances are created by the *Situation Identification* module in Figure 6.3 by considering the value of the `Position` attribute for the `VehicleObject` instances and the `Dimension` and `Position` for the `AreaObject` instances.

The above described relations, relevant for the situation, are evaluated by software agents which acts independently from the current user's goal, thus automatically supporting the data-driven information processing. For instance, let us consider an agent (namely *PM10MonitoringAgent*) which continuously monitors the current level of particulate matter $<$ 10 micrometers (PM10) of an air quality monitoring station in order to send an alarm when the level is greater than a threshold, thus recalling the user attention on an important event happened in the environment which may also lead to the change of the user active goal. The behavior of the agent is reported in Algorithm 6.9. The Sparql[2] query (lines 2-6) retrieves the current value of the PM10 for the sensor whose

[2]https://www.w3.org/TR/sparql11-overview/

URI is passed as a parameter to the agent. The sensing data is described according to the SSN Ontology [238]. Then, the agent monitors continuously (at specific time interval, see line 13) this value by performing the query on the sensor data (line 8). If the PM10 value is too high, an alarm is sent to the application layer as a new event (line 9-12).

Figure 6.9 PM10 Monitoring Agent

```
 1: procedure PM10MonitoringAgent(SensingDeviceUri sd)
 2:     SparqlQuery q =”?PM10Value where
 3:         ?PM10Output ssn:hasValue ?PM10Value.
 4:         ?PM10Output ssn:isProducedBy ssn:SensingDevice.
 5:         ?sensingDevice a ssn:SensingDevice.
 6:          FILTER(?sensingDevice = ” + sd +”)”
 7:     while true do
 8:         PM10Value = ExecuteSparqlQuery(q)
 9:         if PM10Value ¿ PM10Threshold then
10:             Event PM10Alarm = CreatePM10AlarmEvent(PM10Value)
11:             SendEvent(PM10Alarm)
12:         end if
13:         wait(timeInterval)
14:     end while
15: end procedure
```

Using such kind of behaviors, it is possible to monitor important events happening in the environment continuously. All these events are managed and monitored by the Adaptive Goal Selection process (see Figure 6.3) to evaluate if the user should consider another goal as the active one due to the actual condition of the environment, and in this case, it recalls the user attention on such a goal. The presence of such events (represented by the class `EventNotice`, as depicted in Figure 6.7) can in turn affect some relations among `SituationObjects`. When a relation is true, it can lead to the identification of a new `Situation`. This situation can be identified by verifying some rules involving objects, relations, and attributes. As an example, when the truth values of the relations are known, it is possible to identify the situations with the following two inference rules, defined in SWRL. The first inference rule identifies the situation `IncreasingPollution` when the area is already polluted (`PollutedAreaRelation`) and there is a traffic jam (`TrafficJamRelation`) in the same area.

$$\begin{aligned}?PollutedAreaRelation(?pa) \wedge TrafficJamRelation(?tj) \\ \wedge AreaObject(?a) \wedge relation(?tuplepa, ?pa) \\ \wedge relation(?tupletj, ?tj) \wedge onObject(?tuplepa, ?a) \\ \wedge onObject(?tupletj, ?a) \wedge Situation(?s) \\ \rightarrow IncreasingPollution(?s)\end{aligned} \tag{6.1}$$

The second inference rule identifies the situation `IncreasingPollution` when there is an already polluted area (`PollutedAreaRelation`) and there is a driver of a vehicle belonging to the monitored fleet which is driving badly in that area `BadDrivingStyleRelation`, contributing to the increase of the pollution.

$$\begin{aligned}?PollutedAreaRelation(?pa) \\ \wedge BadDrivingStyleRelation(?bd) \\ \wedge InRangeOfRelation?ro) \wedge AreaObject(?a) \wedge Driver(?d) \\ \wedge relation(?tuplepa, ?pa) \wedge relation(?tuplebd, ?bd) \\ \wedge relation(?tuplero, ?ro) \wedge onObject(?tuplepa, ?a) \\ \wedge onObject(?tuplebd, ?dr) \wedge onObject(?tuplero, ?dr) \\ \wedge onObject(?tuplero, ?a) \wedge Situation(?s) \\ \rightarrow IncreasingPollution(?s)\end{aligned} \tag{6.2}$$

6.3 Adaptive Goal Selection

In this section, we describe the computational approach for adaptive goal selection that sustain the human operators in switching coherently between different goals, depending on the current situation, on the salient information, on alarms and conditions, to support users in the alternation between goal-driven and data-driven processing. The approach adopts *desirability* measures for goals in order to evaluate their relevance without the users' intervention. The approach is adaptive as it adapts the process for selecting the goal according to the users' feedback. This is realized by means of a reinforcement learning technique.

6.3.1 Goal Selection Process

The AGSM approach suggest to the human operator the most suitable goal by means of the selection process depicted in Figure

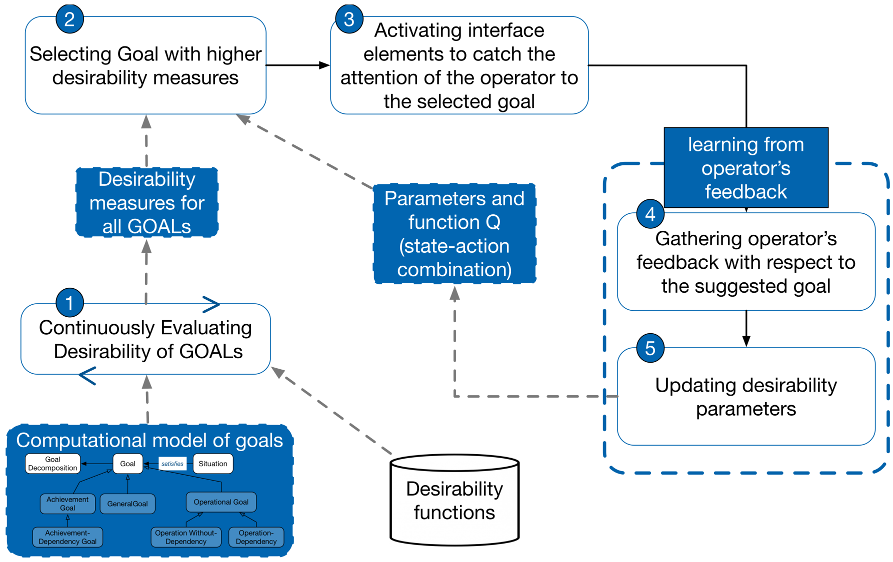


Figure 6.10 Adaptive Goal Selection

6.10.

Let $G = \{g_1, ..., g_k\}$ be the set of all the goals modeled in the Cognition layer. We indicate with $\overline{g} \in G$ the active goal of the user. The aim of the goal selection process is to identify, at each discrete time step $t = t_1, t_2, t_3, ...$, a goal $g \in G$ to suggest to the human operator. The human operator may decide to accept the suggestion and in this case the active goal will become $\overline{g} = g$. The process continuously evaluates the *desirability* of each goal modeled in the Cognition layer (first step in Figure 6.10). Desirability, which will be described more formally in Section 6.3.2, gives an expert-based measure of how much a goal is important for the user, at a given time, by means of a desirability function $d(g) \in [0, 1], \forall g \in G$.

The values of the desirability measures of goals are used to select a goal (second step), by considering both the most desirable goal according to the desirability functions, i.e. the goal $g_i | d(g_i) > d(g_j), \forall i \neq j, i \in \{1..k\}, j \in \{1..k\}$, and the feedback of the human operator. Indeed, the process of goal selection takes into account the reactions of the human operator to the previous goal suggestions, in order to adapt the process itself to the preferences of the user. Specifically, the goal to suggest is chosen with the function $\Phi(D, Q)$ where $D = \{g^1, g^2, g^3, ..., g^k\}, g^i \in G, i = (1, k)$ is the sequence of ordered goal according to their current desirability values, i.e. $d(g^i) \geq d(g^j), \forall i \neq j, i \in \{1..k\}, j \in \{1..k\}$, and Q is a function that takes into account the user's feedback about the past goal suggestions. Details about the Φ function and the Q function are reported in Section 6.3.4.

The goal selected by means of the Φ function is suggested to the human operator by means of special cues on the interface to catch the attention of the user on this goal (third step). Notice that, even if the process selects the most desirable goal, it does not update or change completely the interface. This is because a complete and automatic reconfiguration of the user interface is not useful and it may be harmful causing a critical loss of the SA [15]. This step is executed only if the new selected goal is different from the active one.

Once stimulated by the interface, the human operator can decide to switch to the suggested goal (that, in turn, becomes the new active goal) or to continue to focus the attention on the current active goal. The human operator's behavior represents a feedback regarding the suggested goal (fourth step). The above feedback is gathered and used to update the function Q (fifth step). In brief, the system learns how to suggest goals by exploring how the operators react to suggested goals in specific states. The iterative process (used to calculate Φ), which is executed at periodical intervals of time, combines the top-down mechanism proposing suitable goals by GDTA and human operators' feedback, and the bottom-up mechanism that, for each goal, calculates a desirability measure taking into account contextual and environmental information gathered by sensor data.

6.3.2 Goal Desirability

Goal desirability is an expert-based measure to select the goals in a given context and domain. The value of this measure is computed by means of a desirability function (defined by the domain-experts) applied on the data gathered by the sensors, as proposed in [259]. The desirability of a goal can be influenced by: users' actions, data-driven events (e.g., alarms), identified situations, users' feedback about suggested goals. Other ways to compute the desirability measure can be defined as, for instance, using specific algorithms, heuristics, fuzzy logic. The advantage of using expert-based mathematical function relies in the precision and reliability of the obtained measure which leverages on expert knowledge.

In order to explain the way by which the desirability is computed, let us consider again the set of goals of Figure 6.4. In this example, a logistic operator manages a fleet of vehicles in order to optimize their routes, increment their efficiency, reduce the environmental pollutions. In Figure 6.4, there is one high level goal and three competitive operational goals. Such goals are competitive since the achievement of one goal may be harmful for the others. Consider that, for instance, in order to reduce the delivery time is not always possible to reduce the fatigue of the drivers or to increase the vehicles' efficiency. In these cases, generally the human operator adopts one strategy according to the boundary conditions and to the customers' requests. But, during the delivery, some conditions may change, thus making a different task more useful for optimizing the fleet; nonetheless, the operator may continue to maintain his/her current strategy and to pursue his/her current active goal.

Eq. (6.3) shows the desirability function of goal 1.1. "*Reducing delivery time*"

$$d(g_{1.1}) = \begin{cases} \frac{DeliveryPriority \cdot e^{Delay}}{e^{\alpha}} & \text{if } Delay < \alpha \\ DeliveryPriority & \text{if } Delay \geq \alpha \end{cases} \tag{6.3}$$

where $d(g_{1.1})$ is the desirability of goal 1.1; $DeliveryPriority$ is an index ($0 \leq DeliveryPriority \leq 1$) of the priority of the delivery as arranged with the customer who commissioned the shipping, $Delay = EstimatedArrivalTime - ExpectedDeliveryTime$ is the delay between the estimated arrival time of the shipping and the expected delivery time by the client (when it is negative, it means that the truck is in advance with respect to the delivery time). α represents a threshold of the delay after which the desirability of the goal is at its maximum value.

Another example of a desirability function is shown in Eq. (6.4) that refers to the desirability of goal 1.2 "*Monitor and reduce drivers' fatigue*":

$$d(g_{1.2}) = k \begin{cases} 0 & \text{if } timeSinceLastStop \leq \beta \\ \frac{timeSinceLastStop}{\beta} - 0.5 & \text{if } \beta \leq timeSinceLastStop \geq 1.5\beta \\ 1 & \text{if } timeSinceLastStop \geq 1.5\beta \end{cases} \tag{6.4}$$

where $d(goal_{1.2})$ is the desirability of goal 1.2, k is a normalization factor that depends on the relative importance of this goal with respect to the others; $timeSinceLastStop = CurrentTime - LastStopTime$ represents the time passed since the last stop of the driver; β represents the estimated next stop time and can be set by the operator (e.g., according to the applicable law and regulations).

In order to clarify the evaluation of the desirability of the goal, let us consider the following example that refers to the two above mentioned desirability functions. Figure 6.11 shows the two functions with the following parameters: $DeliveryPriority = 0.24$, $\alpha = 2$, $\beta = 2$, $k = 0.3$. The x axis represents the time, which for the $d(g_{1.1})$ function it is the $Delay$, while for the $d(g_{1.2})$ function it represents the $TimeSinceLastStop$. In this example, the active goal $\overline{g}$ is goal $g_{1.1}$. Over the time, when $timeSinceLastStop$ is equal to 2.8, $d(g_{1.2}) > d(g_{1.1})$ and so goal 1.2 is more desirable than the active one. Considering these values, and the feedback of the users represented by the function Q, the goal is suggested to the user. If the user accepts such a suggestion, the new active goal is executed as described in the next section.

6.3.3 Execution of the Active Goal

When a new goal becomes active, the computational entities (e.g., software agents) responsible for the data processing, situation identification, situation exploitation, and information visualization,

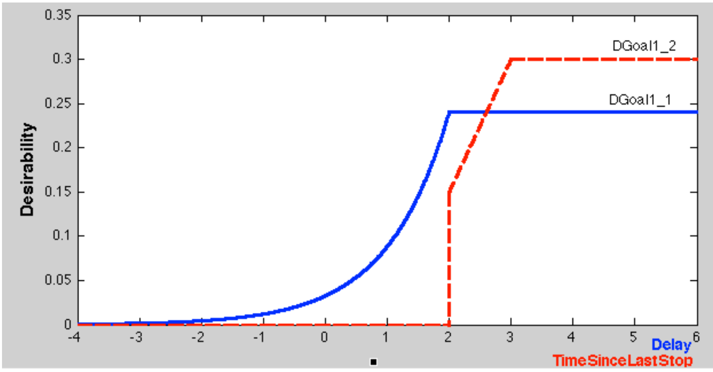


Figure 6.11 Goal Desirability for the fleet management example

have to change their behavior in order to act properly with respect to the new active goal, thus correctly processing the information the user needs. In particular, the semantic representation of the new active goal is copied and instantiated with the necessary information into the Working Memory (WM) (Figure 6.3). The goal is instantiated in a data structure that is graphically represented in Figure 6.12. Such structure, contained in the WM, represents all the information required by the other layers of the approach in order to complete their tasks. Specifically, the figure refers to the instantiation of an example goal "*Evaluating pollution in a given area* ".

When this goal becomes active, the Goal Configurator module of the approach depicted in Figure 6.3 is responsible for updating the Working Memory with the information related to the execution of such goal, by creating the data structure depicted in Figure 6.12. The behavior of the Goal Configurator is briefly described in the pseudocode of Algorithm 6.13. The active goal selected by the user is communicated via an event (line 2), from which the Goal Configurator extracts the reference to the goal representation in the knowledge base (i.e., the URI of the instance of the class representing the goal (subclass of `gso:Goal`). Using this reference, it retrieves the information characterizing the goal by executing a SPARQL query (lines 3-8). As a result, it obtains all the tasks and services that are needed to sustain the active goal. The Goal

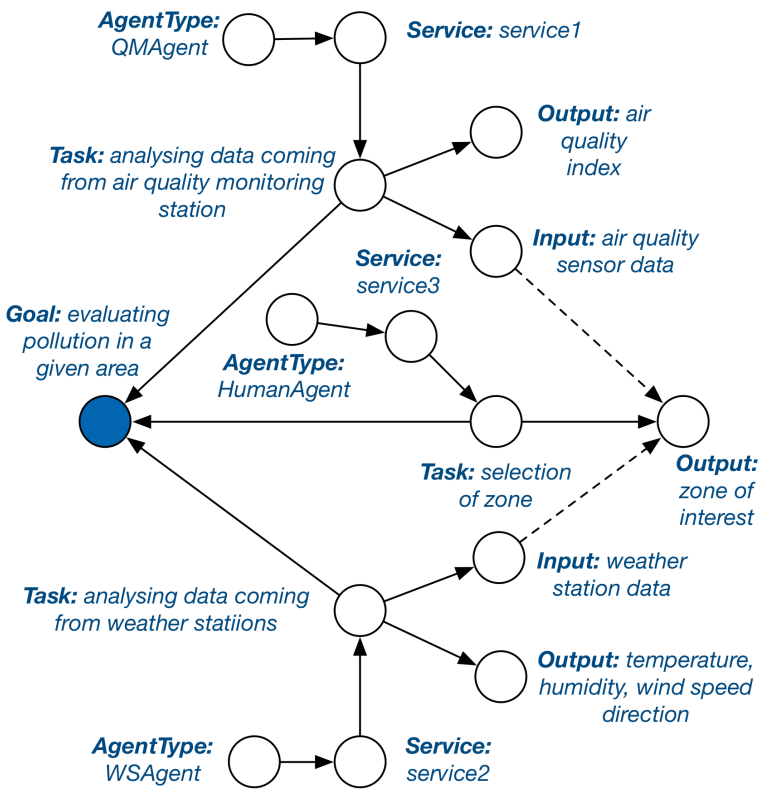


Figure 6.12 Instance modeling of a sample goal.

Configurator updates the Working Memory with the information on the tasks and services by creating an event for each task, which announces the creation of a new computational entity capable of performing such task (line 9-12). In the considered example, the Goal Configurator instantiates three tasks: "*Analyzing data coming from air quality monitoring station*", "*Analyzing data coming from weather stations*" and "*Selection of zone*" as depicted in Figure 6.12, which are needed in order to achieve the active goal. For each task, it instantiates the related services and the agents capable of implementing it. Each task produces some output data (e.g., air quality index, weather station data) that will be added to the Working Memory and made available to other agents (the details are depicted in Figure 6.12).

Figure 6.13 Goal Configurator behaviour

```
1: procedure CoordinatorAgentBehaviour(Event newGoal)
2:     GoalURI gu = parseEventGoal(newGoal)
3:     SparqlQuery q = "Select ?task where
4:         ?task a gso:Task.
5:         ?task gso:support ?goal.
6:         ?goal a gso:Goal.
7:          FILTER(?goal = " + gu +")"
8:     Task[] tasks = ExecuteSparqlQueryOnTheWM(q)
9:     for all Task t ∈ tasks do
10:        Event newTask = CreateEventForTask(t)
11:        SendEvent(newTask)
12:     end for
13: end procedure
```

6.3.4 A Reinforcement Learning Approach to define Φ

In Section 6.3.1, the function $\Phi(D, Q)$ is introduced to choose, at each iteration, the goal that will be recommended to the human operator by considering goal desirability and the human operator's feedback. The function Φ can be defined by using a *reinforcement learning* algorithm [260]. The idea is to gather human operators' feedback regarding past goal suggestions to understand how much a goal, selected and recommended in a given context (state), has been accepted by the human operator and used for the task to execute. In such a way, it is possible to capture the human operator's feedback about the suggested goal for adapting the desirability measures (that have been defined by experts in order to be applied in different contexts, for instance for the logistics operations in different ports) to the specific context (for instance, to the logistics operations in the specific port wherein the human operator acts).

The aim of the reinforcement learning technique is to learn the mapping between the contexts and the actions for maximizing a numerical reward signal. The learning problem is modeled as follows. The environment in which the learning agent acts is the combination of all the goals modeled in the Cognition layer. The action performed by the learning agent is the recommendation of a goal to the human operator. In turn, the human operator provides the learning agent with a reward that corresponds to a feedback on the aforementioned suggested goal. The reward is positive if the feedback is positive. The agent tries to maximize the received reward over time. Specifically, the agent interacts with the environment along a sequence of discrete time steps, $t = 0, 1, 2, 3, ...$. At each time step, the agent receives a representation of the environment state, $S_t \in S$, where S is the set of possible states. $S_t = \langle \overline{g}, g_1, g_2, \ldots, g_k \rangle$ is the state of the environment at time t, where $\overline{g}$ is the active goal for the user/operator, $g_1, g_2, \ldots, g_k$ is the sequence of all plausible goals such that $d\left(g_i\right) \geq d\left(g_{i+1}\right), \forall i = 1, \ldots, k-1$. Function $d(g)$ returns the current desirability measure of goal g by using the proper desirability function (see Section 6.3.2). On the basis of S_t, the agent selects an action $A_t \in A\left(S_t\right)$, $A\left(S_t\right) = \{a \in A | a = suggest(g), (g = \overline{g}) \vee (d\left(g\right) > \rho)\}$, where $a = suggest(g)$ means that the action a consists in recommending the goal g to the user/operator. One time step later, in part as a consequence of its action, the agent receives a numerical reward, $R_{t+1} \in R$, and transits into a new state, S_{t+1}. The reward R_{t+1} is calculated by considering in which extent the human operator has accepted the system suggestion. A simple but effective reward function is: 1 if the suggestion has been accepted, otherwise 0.

At each time step, the agent implements a mapping from states to probabilities of selecting each possible action. This mapping is the agent's policy and is denoted with π_t, where $\pi_t(a|s)$ is the probability that $A_t = a$ if $S_t = s$. In order to specify how the agent changes its policy as a result of its experience it is possible to adopt the State-Action-Reward-State-Action (SARSA) algorithm [261] essentially based on the following update rule:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \left[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)\right] \tag{6.5}$$

In Eq. 6.5, $Q(S_t, A_t)$ is the action-value function, γ is the discount factor and α is is a constant step-size parameter. The used algorithm trades off exploration (suggesting a random goal selected by the vector of goals with high desirability values) and exploitation (suggesting the goal with max Q value in the given state) by executing ϵ-greedy selection [261] on the set $A\left(S_t\right)$. The idea is to run the SARSA algorithm during the system execution and provide its results as suggestions for the human operators. Thus, the selection function Φ is the execution of the SARSA algorithm, whose ability to give suggestions improves along the timeline.

6.4 Evaluation

This section describes the evaluation of the AGSM approach. The objective of the evaluation is to quantify the improvement in term of SA of the operators that is produced by the proposed approach. Specifically, we aim at evaluating if the suggestion of goals, sustaining the alternation of goal-driven and data-driven information processing, is useful for improving SA. The evaluation is performed using the Situation Awareness Global Assessment Technique (SAGAT), described in Section 4.2.1. The AGSM approach has been implemented and evaluated in three prototypical systems in different domains: i) an e-learning system supporting learners in the self-regulation of their activities (Section 6.4.1 ;ii) a fleet management system supporting logistic managers in the control of a fleet of vehicles (Section 6.4.2); iii) a Decision Support System (DSS) for supporting logistic operators in the management of operations for handling containers in a port terminal (Section 6.4.3). Following the SAGAT approach, the experiments for the evaluation is performed in the following way for each system:

1. analyzing the domain and the tasks of the users in order to identify the users goals and the SA information requirements by means of the GDTA technique;

2. defining the simulation scenarios, i.e., which is the task the user should perform and what are the surrounding conditions according with the goals and the SA requirements identified with the GDTA;

3. defining a questionnaire for evaluating the three levels of SA, with respect to the SA information requirements identified with the GDTA technique.

4. executing the experiments by involving the users: each user interacts with the system to complete the simulation scenarios. Each user may execute more trials. The simulations are performed by using two versions of each system: A) the one that does not implement the AGSM approach; B) the one that implements the AGSM approach. The users do not know which version of the system they are using. In this way, we can quantify and measure the improvements in the SA levels given by the AGSM. Some queries, randomly selected from the questionnaire, are presented to the user at random time instants by freezing the simulation. The queries are answered by the users by means of a web application.

5. collecting the answers and evaluating the results in terms of levels of SA gained by the users, comparing the two versions of the system.

Furthermore, to evaluate the usefulness of the AGSM approach in supporting SA and thus improving the decision making performance, we realized a numerical simulation of a port terminal container using the simulation environment Arena®. In this simulation, we suppose that the logistic operator always accepts the suggestions of the AGSM approach regarding the active goals, and that he/she makes the decisions about the handling of containers in the port according to the active goals. In this way, we can quantify the improvement in the performance of the (simulated) port terminal container by measuring some Key Performance Indicators (KPI). This evaluation is described in Section 6.4.3.4.

6.4.1 SA System for supporting Self-regulated Learning

The first evaluation of the AGSM approach is conducted in the domain of e-learning, by implementing a prototype for supporting self-regulated learning. Self-regulated learning is defined as an active and constructive process wherein learners monitor, regulate and organize their cognition, motivation and behavior according to one or more learning goals [262]. The main motivation for the application of AGSM to self-regulated learning comes from the consideration that many learners can have serious difficulties in benefiting from informal and not-formal learning experiences, considering that having a control on the whole learning processes is a complex task for the learners and considering also their lack of ability in self-regulating such processes. In particular, in hypermedia systems, learners have difficulties to choose autonomously the right set and sequence of experiences with respect to a more or less formalized learning goal. We argue that one of the main hindrances for benefiting from Seamless Learning and for efficiently self-regulating learning in such scenarios can be found in the lack of learners' Situation Awareness with respect to their learning processes and in the difficulty in selecting the most suitable learning goal in a given situation. Thus, we evaluate the capability of AGSM approach to increase the situation awareness levels of learners by comparing two versions of the prototype: a first version (A) which does not implement the approach, and a second version (B) which implements the AGSM approach, thus to quantitative measure the increase of the level of SA via the SAGAT approach.

To perform the evaluation, first we need to instantiate the generic ontological model of goals and situation to the specific e-learning domain, as described in next subsection.

6.4.1.1 Extending the Semantic Model to the e-Learning domain

Before we can define a model for representing situations, we need to identify the entities and the relationships relevant for this domain. First of all, the *environments* in which the situations occur consist of conceptualizations of subject matters. More in detail, these conceptualizations are represented by *concept maps*, in which each node represents a specific concept or topic (e.g., if the concept map is related to the *Semantic Web* argument, some concepts could be *RDF*, *ontology*, *OWL*, *ontology-building*) that is linked to other concepts by means of thematic relations (e.g., OWL concept has a thematic relation with ontology concept), hierarchical or part-of relations (e.g., *RDF-Serialization-language* concept can be the ancestor of *RDF/XML*, *Turtle*, *N3* concepts) or propaedeutic relations that can represent the suggested order for learning a sequence of concepts (e.g., *RDF* is propaedeutic for *OWL*). Thus, a concept map represents the conceptual environment in which learners and teachers execute their learning activities. In particular, when the learner executes a specific learning activity (e.g., reading an article about *RDF Specification* on an e-learning platform), he/she acquires knowledge about the involved topic (e.g., *RDF*) and this, as a side effect, produces changes in the controlled environment represented by the concept map. Conceptually, these changes represent traces (left by the learners on the map) which can be gathered and analyzed in order to understand the situations in which the learners are. Indeed, the process of situation identification in this case means to identify the way by which learners perform their activities. Regarding the *situation*, it can be seen, at the highest level of abstraction, as the *learning state* in a spe-

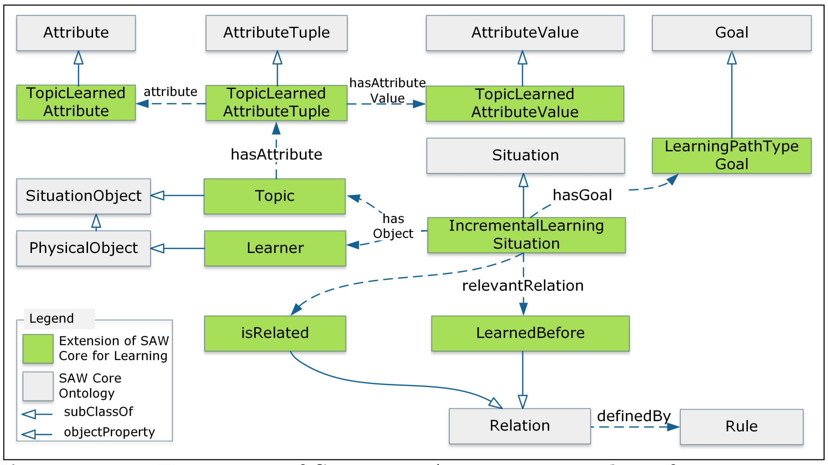


Figure 6.14 Extension of Situation Awareness ontology for representing Incremental Learning Situation

cific time slice of one or more learners. Learners, in fact, need to know their learning state in order to be able to self-regulate their learning process. This learning state can be represented by different elements, like the learning path type (i.e., the way learners acquire knowledge), the number of concepts already acquired, the percentage of achievement of a learning objective, and so on. These elements can be monitored and evaluated by observing the traces left by the learners on the map.

Having described the main elements of the seamless learning scenarios in which we are interested, we can define the model for representing situations.

The situation model specializes the semantic model described in Section 6.2.3 to the Seamless Learning domain.

Figure 6.14 depicts the proposed model by means of an example consisting of one situation and three relations. In the figure, the white rectangles show the main classes of the SAW core ontology, whilst the classes of our proposed model are depicted in green boxes. Notice that our classes specialize (by means of the `rdfs:subClass` property) the classes of the SAW core ontology. Let us consider the following example scenario in order to explain the extension of the SAW Core ontology for seamless learning.

In order to self-regulate the learning process, firstly learners need to be aware of and to monitor the way in which they acquire knowledge (i.e., the way by which they choose their learning activities, the order in which they acquire new concepts, and so on), which can be represented in terms of *learning path type*. We represent the learning path type as one of the `Goal` that the learners would monitor. In figure 6.14, the `LearningPathTypeGoal` is a subclass of the class Goal. The different learning path types that the learner could exhibit are represented as subclasses of `Situations`, related to the `Goal` by means of `hasGoal` object property. Let us focus on the `IncrementalLearningSituation`: this situation represents the behavior of a learner which acquires concepts that are directly linked by means of propaedeutic or thematic relations. This learning path type is the counterpart of the `SparseLearningSituation`, in which learners acquire new concepts that are not linked to already learned ones.

In order to identify if the *IncrementalLearningSituation* satisfies the goal, for each new learning activities executed by the learner, we need to verify if the following relations hold: i) `IsRelated`, in order to verify that the new acquired concept C_2 has a relation with another concept C_1; ii) `LearnedBefore`, in order to verify if the learner have acquired the concept C_1 before the concept C_2.

In order to evaluate if the `IncrementalLearningSituation` holds, two kinds of `SituationObject` are considered: `Topic`, whose instances represent the concepts of the concept map, and `Learner`. The `TopicLearnedAttribute` for representing if a topic has been already learned; it is used for inferring if the `LearnedBefore` and `IsRelated` relations hold.

Figure 6.15 depicts the subclasses representing the events and the rules that allows verifying the aforementioned relations and thus if the situation `IncrementalLearningSituation` holds. In this case, one of the event that have to be monitored is the acquisition of a new topic (of the concept map) by the learner. This event causes the change of the value for the attribute *TopicLearnedAttributeValue* and the firing of the rule *LearnedBeforeRule*.

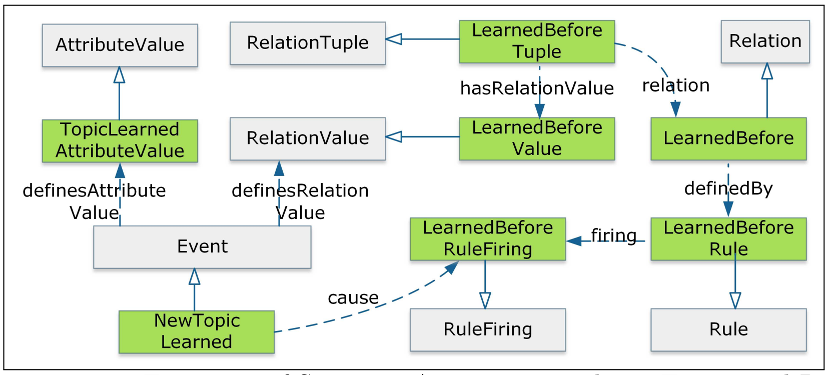


Figure 6.15 Extension of Situation Awareness ontology: Events and Rules

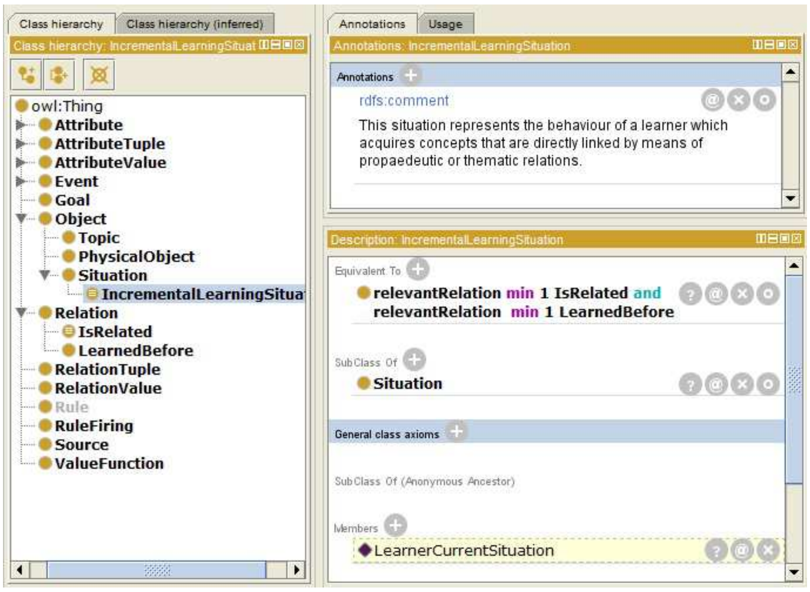


Figure 6.16 Situation Identification using OWL reasoning

The rule for identifying *IncrementalPathTypeSituation*, given that the *Relations*, supporting this situation, are already known, is depicted in Figure 6.16, implemented in Protégé 5.0[3]. Specifically, when a new `Event` of type *NewTopicLearned* occurs, it causes the firing of the rules related the two aforementioned relations *IsRelated* and *LearnedBefore*. In this case, two instances of the relations are added to the ABOX (extensional knowledge of the knowledge base) of the Working Memory (WM). Assuming that these two relations hold and the corresponding facts have been already added to the WM, one can verify that the user is involved in the situation *IncrementalLearningSituation* by means of the rule described in Figure 6.16. In particular, a generic *Situation* can be classified as an *IncrementalLearningSituation* if, at a certain time instant t, exist an instance of the *Relation IsRelated* and an instance of *LearnedBefore* involving the same instances of *Learner* and *Topic* which are related to that situation by the property *hasObject*.

6.4.1.2 Prototype for Self-regulated Learning

The approach described in the previous section has been implemented in a software prototype in the context of Social Learning [263]. It provides social learning environments in which learners browse contents and participate in discussions with other learners. In such a context, we want to provide the learners with a tool, integrated in the social learning platform, which helps them to understand how their learning processes are evolving, so to support them in choosing the best learning activities for fulfilling their objectives.

Figure 6.17 provides a sketch of the interface of the proposed prototype. The interface consists of a main area (on the left) in which learners browse contents, read articles, share opinions with other learners, provide comments and so on. On the right, instead, there is a graphical representation of the concept map, which shows through different colors, the already learned concepts,

[3] Protégé http://protege.stanford.edu/

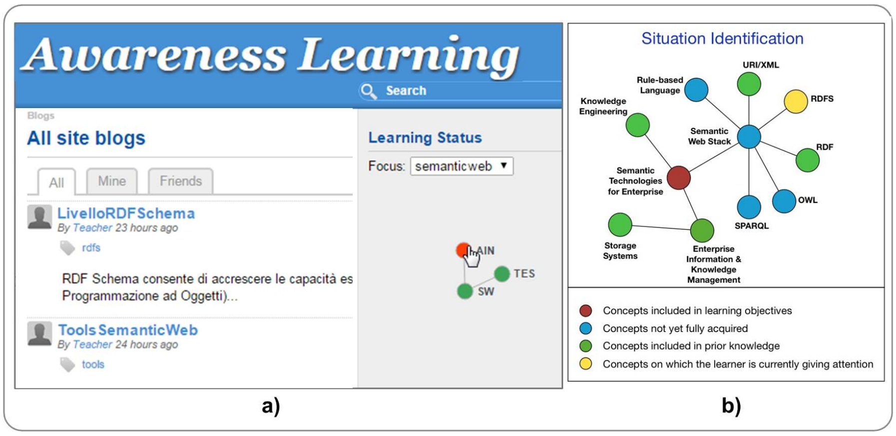


Figure 6.17 Screenshot of the prototype

the ones belonging to the current objective, the concepts that are needed for fulfilling the objective, and so on. This graphical representation of the map shows only the portion of the map that is relevant for the current activity of the learner. The learner can also see the overall map, in order to have a global view on his/her current learning progress. The situation identified with the approaches described in previous sections are directly shown via the map. The map is used also for the suggestions about the change of the active goal.

In order to increase the learners' situation awareness, the interface has been designed in accordance with some of the principles of designing for SA proposed by Endsley and Jones in their book [15]. The aim is to provide learners with the needed information as quickly as possible and without too much cognitive effort. Table 6.1 summarizes the main principles for designing situation aware interfaces, highlighting the main elements of the proposed prototype that follow these principles.

6.4.1.3 Data and Method

To conduct the experiment, we simulated three scenarios based on the requirements we used to design the system as foreseen by the methodology. In such scenarios, the considered learning situa-

Table 6.1 Elements of the prototype that satisfy the principles of designing for SA

N.	Principle	Prototype
1	**Organize information around goals.** Presenting information in terms of operators' goals rather than in a technology-oriented way	The goal of the user is to understand his/her learning progress. The map clearly show this by depicting current activities and the best path for achieving the goal
2	**Support comprehension.** Present Level 2 information directly in order to reduce operators' cognitive workload	Examples: 1) Users can directly see which are the concepts linked to current learning activities or objectives, so they do not need to search for such information. 2) Users can see the percentage of objective completion, without the need to calculate it.
3	**Provide assistance for SA projections.**	Users can easily project their learning situations in near future because the prototype shows their learning path type and the already acquired concept. Thanks to this information, they can self-regulate their learning style being able to identify the weakness in their learning progress
4	**Support global SA.** Provide users with the "big picture": a high level overview of the situation across operator goals	The concept map can be zoomed in and out in order to provide both the global picture and a more detailed view on particular concept.
5	**Support trade-offs between goal-driven and data-driven processing.** Take into account both top-down (i.e., Principle 1) and bottom-up processing (i.e., Principle 4)	The detailed view of the map showed when the user is reading an article or executing other activities on the platform, follows the goal-driven processing of data, because it focuses the attention of the user on its current goal. The global view of the map support global SA with a data-driven processing because it directs the user in order to focus his/her attention to particular portion of the global concept map, in order to achieve high-priority goals.
6	**Make critical cues for schema activation salient.** The critical cues for supporting decision-making need to be made salient in the interface design.	Critical cues are highlighted by means of different colors on the map. For instance, different colors are used in order to indicate the concepts that the user needs to know in order to fulfil an objective.
7	**Take advantage of parallel processing capabilities.** Use multi-modalities interface in order to limit information overload.	The concept map is displayed as a simple graphical interface, so this principle is not applicable to it. The social platform, instead, supports multi-modalities by means of multimedia contents (e.g., video, audio, graphics)
8	**Use information filtering carefully.** Avoid computer-driving strategies for filtering information but provides the operators with tools for determining what they will look at when.	The interface gives to users the possibility to filter the concept map by focusing the map on a specific concept (which become the center of the map) or by filtering and eliminating some concepts from the map (e.g., to eliminate some activities, to show only the objectives, to eliminate already acquired concepts).

tions consist of different learning objectives (at least one learning objectives is active) and different sets of already learned concept (also partially activated concepts). Each concept map of such scenario has its own complexity in terms of number of concepts and relations among them. The first scenario mainly represents an incremental learning situation, with one main current objective; the overall map consists of 20 concepts. The second one, instead, represent a sparse learning situation, with already learned concepts not directly linked among them and with one main current objective; the map is slightly different form the previous one but consists again of 20 concepts. The last scenario is trickier. The map contains 35 concepts; some of the already learned concepts have some direct connections among them and with the final objective, while other are sparse concepts, thus to hinder the identification of the learning situation to the user.

For each scenario, the elements the user has to perceive at level 1 relates to the identification of the already learned concepts and their relations, the current objective, the current learning activity he or she is performing and the related learning concept. At level 2, the comprehension is assessed by asking more complex questions that try to understand if the user is able to identify the relations among the concepts needed to fulfill the objective and the once he/she already knows. Questions for level 3 aim at assessing if the user is able to make decisions about the best activities he/she can do according to his/her current objective and which can be the next learning objective to achieve. For questions of level 3 we also ask the motivations for the given answers. For the quantitative answers (e.g., *how many concepts do you already known on this subject matter?*) we permit a range of $\pm 20\%$ around the actual answer. Table 6.2 shows the entire questionnaire from which we extract the questions to submit to the users during the freeze of the simulation. For each question, we define the correct answers in order to define a ground truth for evaluating users' answers. The last column of table four indicate the SA Level to which the question mainly refers to. The simulation tool we have defined is able to select just those questions that are significant according

Table 6.2 Questionnaire for SAGAT evaluation.

ID	Query	SA Level
1	How many concepts of the map you have already learned?	1
2	Which is the last concept you have learned (also partially)	1
3	Which is the current objective?	1
4	Which is the concept you are learning now?	1
5	Which are the suggested activities to do?	1
6	According to the map, are there some concepts you have not totally understand/learned	2
7	Are the majority of the concepts you have learned so far directly connected by some arcs?	2
8	Does it exist a sequence of concepts that connect the concepts you have learned and your current objective?	2
9	According to you, which could be the next concept it is convenient to learn for reaching the current objective?	3
10	Which of the following activities do you want to do next?	3
11	Which can be the next learning objective?	3

to the state of the simulation at the frozen point. In order to avoid biases and to avoid that the users can put their attention on particular aspects of the simulation (due to influences of previous questions), the order of the questions and the frozen time instants are selected randomly. Lastly, each user performs the simulation by himself/herself, in order to avoid collaboration and interference among them.

The sample consists of 15 students of bachelor's and master's degrees in Computer Engineering and Management Engineering at the University of Salerno (Italy). Accordingly, considering the 15 students and the 3 scenarios and the two modalities of execution of the system (i.e., with and without the AGSM approach), the experiments have been conducted in order to guarantee that each item of the questionnaire has a number of answers between 30 and 45 (usually, it is recommended that each query has between 30 and 60 sampling for each design options of the system, as discussed in Section 4.2.1).

Before the simulation starts, we explain the system to the students and let them to familiarize with the interface. We also

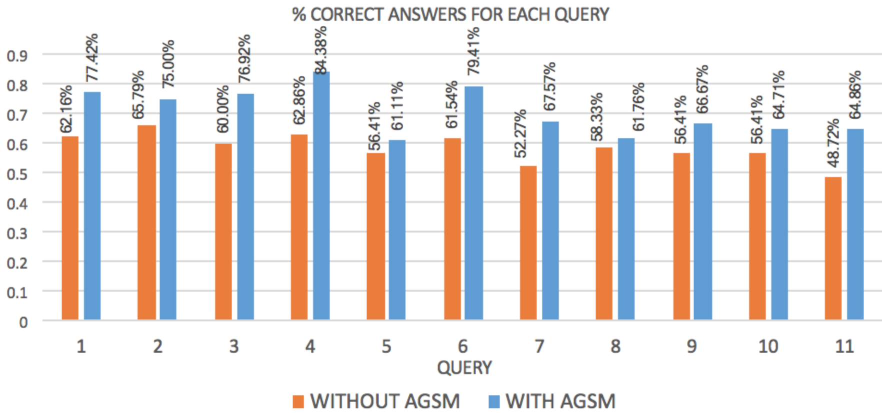


Figure 6.18 Evaluation results: percentages of correct answers for each query.

explain the SAGAT procedures and the way for answering each query via the form (let them performing a couple of trials with the system), in order to avoid misunderstandings regarding both the interface and the questionnaire that may alter the measurements of SA levels [264].

6.4.1.4 Evaluation Results and Discussion

Figure 6.18 shows the percentage of correct answers given by the users for each of the 11 queries of the questionnaire, as resulting by the experiments, by comparing the two versions of the system: without the AGSM (in orange) and with AGSM (in blue).

The first 5 queries of the questionnaire are mainly devoted at measuring SA Level 1. It is possible to observe a good improvement at this level. This is mainly due to the fact that, thanks to the information clearly shown by the concept map, it is quite simple to perceive basic elements like the number of concepts the user already knows, which concept represents the objective, and so on. Notice that the lowest value is that of query number 5, asking which are the suggested activities to do. This low percentage suggests us to modify the way by which we currently propose the activities to the user. Currently, the users may identify the activities they can do by clicking on a specific concept. This suggests to modify the system in order to provide the users with direct notifications about the new activities that are congruent with their learning situations and objectives.

Queries 6,7 and 8 are mainly devoted at measuring Level 2 SA. The improvements for query 6 are greater than the other two. This is because the user can easily be aware of the concepts he/she does not completely know simply by looking at the colors of the concepts in the map. The other two queries in Level 2, instead, require a higher cognitive workload and a good working memory, as the user needs to observe and remember if there are some paths on the graph (concept map) that connect his/her current activity to the final objective.

Considering SA Level 3, represented by queries 9-11, we observe an improvement with respect to the version of the system without AGSM, but this improvement is lower than that of the other two levels. This is due to the fact that increasing the projection capabilities of the students is difficult. The capabilities to project the current state in the future and to act according to it is difficult for non-expert users, as it requires the proper mental model and the adequate experience. This improvement can be achieved by assisting the learner in chooses the next activity to do or the next objective to achieve (i.e., by exploiting the prerequisite relations among concepts and the profile of the user in term of competencies and preferences).

Figure 6.19 depicts the percentages of correct answers for each scenario, thus showing the impact of the difficulty of the scenario on the performance of the users. The figure clearly shows the improvement given by the AGSM approach in all the scenarios. As it can be expected, the performance in the first scenario are higher than the other two, due to the fact that this scenario is simpler than the other (a lower number of elements should be perceived in this scenario). However, notice that the improvement with respect to the version of the system without the AGSM is lower than in the other two scenarios. This is because, even without the AGSM approach, it is simpler for the students to

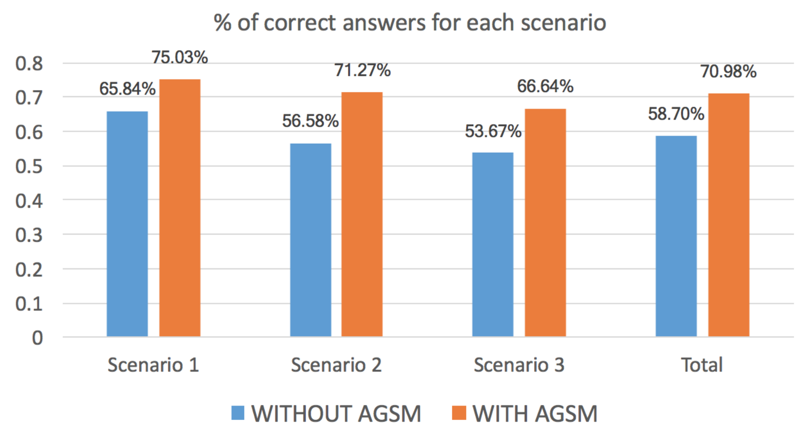


Figure 6.19 Evaluation results: percentages of correct answers for each scenario.

answers correctly to the questionnaire due to the characteristics of the first scenario. Indeed, in scenarios that are more tricky, the percentage improvement is greater, demonstrating the capability of the approach in concretely supporting the students in self-regulating their activities- Concluding, the evaluation shows good results since the AGSM approach allows improving the level of SA of the operators of more than 12%.

6.4.2 A Green Fleet Management System to improve SA

In this section, we describe the evaluation of the AGSM, and specifically its support in the alternation between goal-driven and data-driven information processing, by means of a prototypical system for the management of a fleet of vehicles. Again we use SAGAT to compare two versions of the system (with and without the AGSM approach) in order to quantify the improvement in terms of SA given by the approach.

6.4.2.1 The Green Fleet Management System

In this section we describe how AGSM is instantiated for realizing a novel Green Fleet Management System (GFMS), i.e. a Fleet Management System [265] supporting logistic operators in maintaining high level of SA while controlling the fleet in a sustainable and ecological way, trying to reduce its emissions. The reduction of the emissions mainly depends on two aspects: i) the behavior of the drivers; ii) environmental and contextual factors like, for instance, weather and traffic, which may impact on the current routes of the fleet. Fleet management systems operators and drivers need to work synergistically for obtaining an efficient and eco-friendly fleet. Indeed, logistic planners need tools that simplify their role by assessing the current state of the fleet and planning the best routes (which cannot be statically planned early, but needs to be adapted dynamically following the environmental conditions) while drivers need information on the route they have to follow, on their truck and on their current driving style (i.e., suggestions about what they can do for maintaining an eco-friendly driving style). In order to achieve this, the operators (in the control room) need to consider a huge amount of information and to give attention to different and competing goals. Information of the traffic, the weather, the position of the trucks, the destination of the shipments and the possible paths, the current emissions of the fleet, are just a few of the parameters that the operators should consider. And they have to process this information while performing their usual tasks like planning future shipments, dispatch customers' orders, control the warehouse and the stocks, and so on. As a result, they surely need a tool that reduces this huge cognitive workload and at the same time tries to improve their situation awareness, also with respect to the environmental aspects of their work. To this aim, we have designed and implemented the GFMS.

The architecture of the GFMS, which implements all the functionalities of the AGSM approach (see Section 6.2.1), is based on a multi-agent framework, as depicted in Figure 6.20. The system

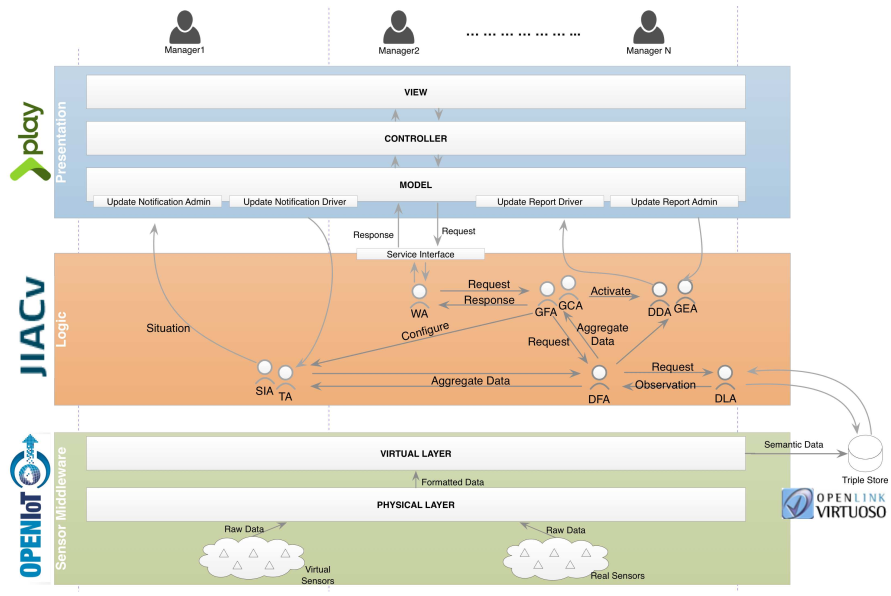


Figure 6.20 Technologies adopted for the realization of GFMS.

consists of three main layers: i) *Sensor Middleware* for the management of sensors and the gathering and memorization of sensor data according to the semantic model; ii) *Logic*, which contains the software agents for processing sensor data, identifying the situations and managing the goals with the AGSM approach; iii) *Presentation*, for the information visualization according to the users' goals. The agents of the Logic layer are implemented by using JIAC V[4], a Java-based open source agents framework. One of the main advantages of the adopted framework is its support to the exposition of web services (whose behaviors are implemented by agents) that eases the interoperability of the prototype with external systems and with the interface layer. Table 6.21 describes the role of the main agents of this layer. Sensor lifecycle, sensor data acquisition and semantic sensor data representation are managed by using OpenIoT[5] middleware infrastructure. This middleware allows to gather sensor data (via the xGSN middleware) and to

[4]http://www.jiac.de
[5]https://github.com/OpenIotOrg/openiot

Figure 6.21 Main agents of the framework

Agent	Behavior
Data Linker Agent (DLA)	It gathers data from the triple store
Data Fusion Agent (DFA)	Starting form data gathered by DLA, it processes and fuses heterogeneous data in order to generate observations (high-level information)
Situation Identification Agent (SIA)	It identifies a situation by processing the observations generated by the DFA agents. Only the situations relevant for the current goal are identified.
Data-driven Agent (DDA)	Such kind of agents processes raw sensor data in order to verify alarm conditions, events, abnormal conditions, even if they are not related with the current goal.
Goal Evaluator Agent (GEA)	It evaluates the desirability of each goal, by using the information provided by the DDA and those contained in the Cognition Layer.
Goal Coordinator Agent (GCA)	It manages all the other agents of the framework. When a new goal is instantiated in the Working Memory, it creates all the Task agents needed for completing the goal. It verifies the realizability of the goal and/or it tries to make the goal realizable.
Goal Configurator Agent (GFA)	Its role is to configure the Working Memory with the goal selected by the user.
Task Agent (TA)	The task agent is a generic agent which performs one of the task needed for satisfying the selected goal.
Web Agent (WA)	It manages the interactions with the Web Application of the Presentation Layer.

represent them with SSN Ontology by using Linked Sensor Middleware (LSM)[6]. Semantic sensor data is then stored into the quad store Virtuoso Open Source Edition[7]. The Presentation Layer and the graphical user interface are realized by using the Play! Framework[8]. It is implemented by means of the Model-View-Controller pattern. It communicates with the agents by means of web services.

The interface of the GFMS is a web application consisting of different views. The design of the user interface is based on the principles of design for situation awareness [15]. The main interface is shown in Fig. 6.22. It is designed around users' goals, following the GDTA approach. The set of main goals supported by the prototype is shown in the hierarchy of Figure 6.4. One of goal of the logistic operator of the GFMS is to increase vehicle efficiency, by reducing emissions and fuel consumption. Specifically, in order to pursue this goal, one of the task of the user is to evaluate the current pollution along the predefined route followed by

[6]https://code.google.com/p/deri-lsm/
[7]http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/
[8]https://www.playframework.com

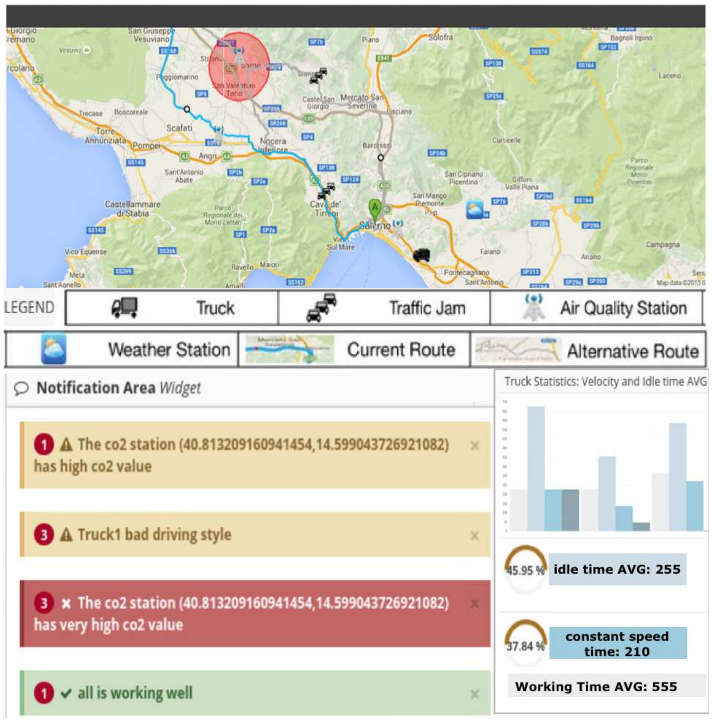


Figure 6.22 Main interface of the Green Fleet Management System

the vehicles (as we have discussed also in the examples of Section 6.2.3. If the level of pollution is too high, the operator may decide to change the current routes of some vehicles or to send feedback to drivers in order to suggest a change in their driving style. The main area of the interface consists of a map showing the current position of vehicles, the position of air quality monitoring stations and weather stations. Moreover, the map shows the current route the driver should follow. The operator may select a specific area (by drawing a circle on the map) in order to evaluate the pollution in that area (see the red circle in Fig. 6.22). The color of the circle gives an immediate insight on the state of the pollution. A red area indicates a high polluted area, yellow for moderate pollution, green for good air quality. At the bottom of the main area, we find two of the available widgets in the prototype. At the right-hand side there is an area devoted to show the driving behavior in terms of number of acceleration, braking and steering, and velocity and idle time of the truck. At the left-hand side of the interface, at the bottom, there is a *Notification Area*. Such area notifies the events which happen in the environment and that requires operator's attention, according to the data-driven information processing approach. Notice that the notifications about the suggestion of a more desirable goal are shown in this tab only in the version of the system which implements the AGSM approach. Figure 6.23 shows another view of the GFMS which allows users to analyze reports and statistics about the pollutant emissions at different time intervals, and statistics about drivers' behaviors. Such detailed information are useful for pursuing some of the goals of the GDTA, as for monitoring drivers' fatigue and evaluating pollution.

6.4.2.2 Data and Method

SAGAT is adopted to assess the SA improvement thanks to the implementation of the AGSM approach in the prototypical system. To conduct the experiment, we simulated three scenarios summarized in Table 6.3. The main conditions that should be understood

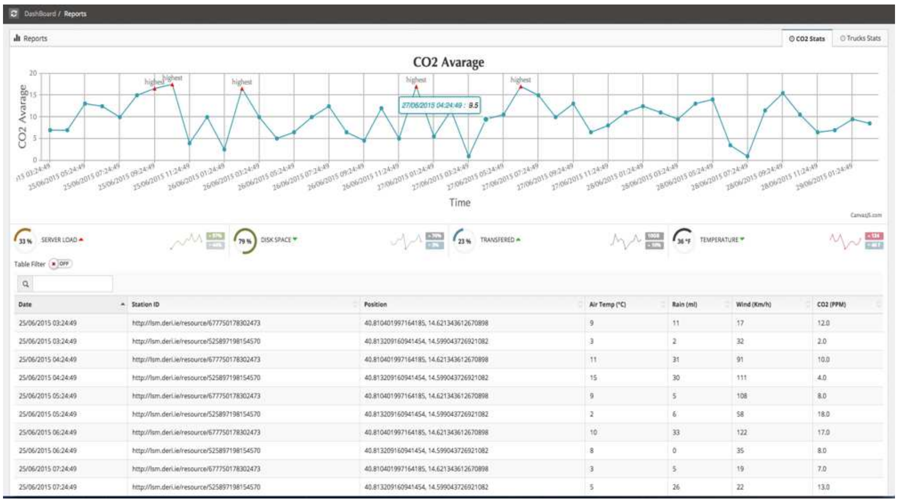


Figure 6.23 Interface of the GFMS showing reports and stats.

by the users in order to make the correct decisions are reported in the table for each scenario. These scenarios have been defined in the context of the Research&Development project (funded by the Italian Ministry of Instruction, University and Research) Mar.Te[9]. One of the aims of this project is to reconfigure the sea-land logistic processes (especially in the Salerno-Naples area, in Italy) and to define practical solutions to rationalize the most critical management and organizational processes of port and freight terminal logistics activities, even in order to reduce the environmental footprint of the sea-land logistics.

In these scenarios, all the vehicles are fully loaded. An alternative route is a path that can bring the truck at the same destination. Vehicles are distributed in the alternative routes, while more than one vehicle can be on a same route. In each scenario, one truck of the fleet may be the main source of an increase of pollution because of the driver's guiding style and/or the traffic jam of the route. For each scenario, the elements to perceive depends on the information requirements identified as a result of the GDTA approach. At level 1 SA, the elements to perceive relate to the

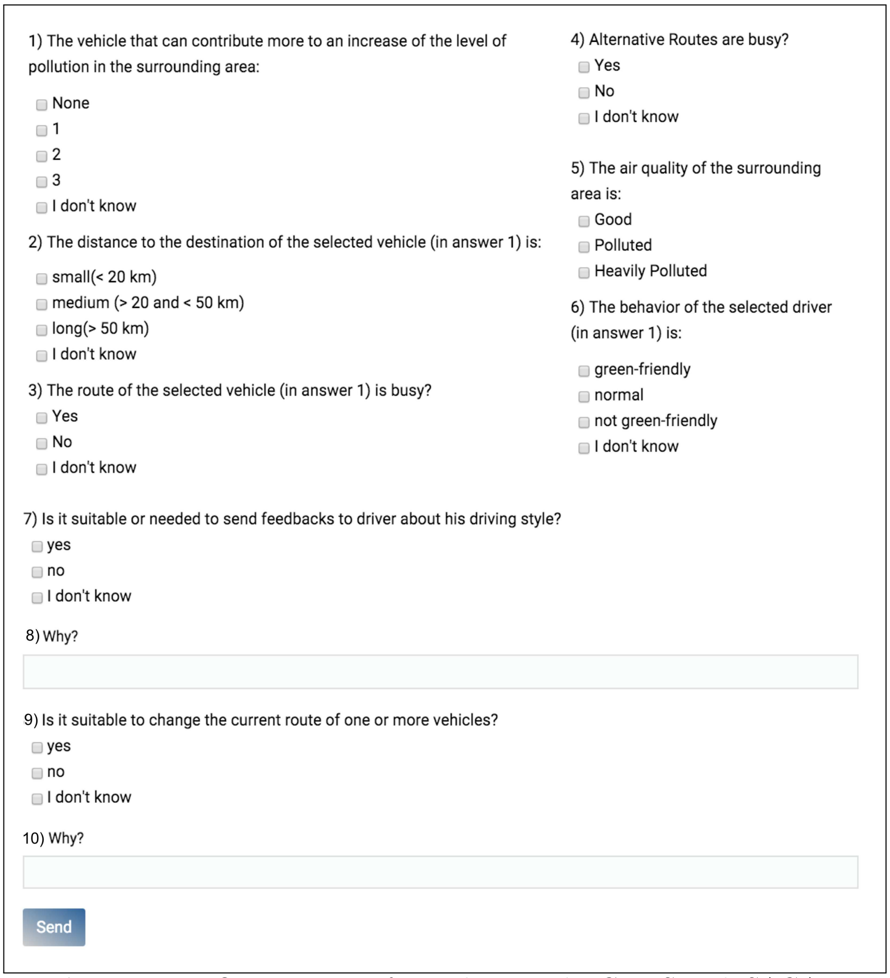
[9]http://mar-te.com and http://www.corisa.it/project/martem/

Table 6.3 Evaluation scenarios

Id	# of vehicles	# of routes	Conditions	Event supplied (in the case of AGSM approach)
1	3	5	Truck n. 3 is a potential source of pollution because of aggressive styling guide of the driver	Alarm on the guiding styles of the drivers
2	3	5	Truck n. 3 is a potential source of pollution because of both traffic situation on its route, and aggressive guiding style of the driver. Alternative routes are busy.	Info on the traffic situation of the possible routes
3	6	3	Truck n. 3 is a potential source of pollution because of traffic situation on its route and aggressive guiding style of the driver. One alternative route is busy. The other alternative route is longer.	Alarm on the guiding styles of the drivers; Info on the traffic situation of the possible routes; Info on the pollution level of the areas.

identification of the specific vehicle that can become dangerous for the pollution on its way to the destination. The comprehension at level 2 SA is assessed via questions on the traffic situation of the route where there is the dangerous truck, on the level of pollution in the area, on the guiding style of the dangerous truck. At level 3 we ask what is the best decision between giving feedbacks to the driver to i) change his guiding style or ii) proposing an alternative route. We also ask motivations for the choices. Fig 6.24 shows the questionnaire from which we extract the queries to submit to the users. As depicted in the figure, the questionnaire is submitted to the users as a web page during the frozen points of the simulation. This eases the process of answering the queries as well as the process of gathering the data.

Each scenario is executed twice: without and with the supply of the events reported in the last column that represent the special cues used by the AGSM approach to suggest to focus on a different goal. The scenarios are simulated in randomized order. The sample involved in the experiment consists of 15 stakeholders of the MAR.TE. project. This ensure us to obtain a number of answers for each query between 30 and 45, thus respecting the guidelines of SAGAT (see Section 4.2.1). The subjects have been trained for using the GFMS and they have familiarized with the system in different trials. Moreover, we explain the SAGAT procedure to the subjects and the way by which they should answer



1) The vehicle that can contribute more to an increase of the level of pollution in the surrounding area:

- None
- 1
- 2
- 3
- I don't know

2) The distance to the destination of the selected vehicle (in answer 1) is:

- small(< 20 km)
- medium (> 20 and < 50 km)
- long(> 50 km)
- I don't know

3) The route of the selected vehicle (in answer 1) is busy?

- Yes
- No
- I don't know

4) Alternative Routes are busy?

- Yes
- No
- I don't know

5) The air quality of the surrounding area is:

- Good
- Polluted
- Heavily Polluted

6) The behavior of the selected driver (in answer 1) is:

- green-friendly
- normal
- not green-friendly
- I don't know

7) Is it suitable or needed to send feedbacks to driver about his driving style?

- yes
- no
- I don't know

8) Why?

9) Is it suitable to change the current route of one or more vehicles?

- yes
- no
- I don't know

10) Why?

Send

Figure 6.24 Questionnaire for evaluating the GFMS with SAGAT the questions (as discussed in Section 4.2.1).

6.4.2.3 Evaluation Results and Discussion

The results are reported in Figures 6.25.a-f that compare the average percentage of correct answers for the three scenarios, respectively, in the two modalities of execution of GFMS (without and with the AGSM). Specifically, Figures 6.25.a-c show the percentage of correct answers for each of the ten queries of the questionnaire, respectively for Scenario 1, 2 and 3 in the two modalities. Figures 12.d-f show the average of correct answers clustered by the level of Situation Awareness to whose they refer. Also in this case, we compare the two modalities for each scenario.

We observe an average improvement of SA for all the scenarios. SA are improved thanks to the AGSM approach with an average percentage increase of 17.78% of correct answers. Specifically, in the first two scenarios we obtain a clear improvement for all the three levels of SA, while in the third scenario (Figure 6.25.(f)) we observe just a slight increase in correct answers at SA level 2 (Comprehension) but we do not observe a substantial improvement at SA level 1 and SA level 3. If we analyze the specific users' answers, in some cases the provision of alarms and notifications (sent by the AGSM) provoked an unwanted change in the answers at level 3 in the third scenario. We argue this is mainly due to the difficulty of the third scenario (in which the users have to control many trucks in a situation in which many roads are congested). In some cases, analyzing the motivations provided, it seems that the suggestion to change the active goal provided by the AGSM (although correct) has confused the users, leading them to to make a wrong decision. Indeed, some users make error in projecting situations or assessing the current situation on the basis of previous experience (e.g. it is not safe to change to longer route because the traffic may decrease) or also they conclude that gain and losses are correlated (e.g. it is not safe to change because the route is longer and thus the pollution increase) leading to a sort of zero-sum bias. This is due to the lack of correct mental models and

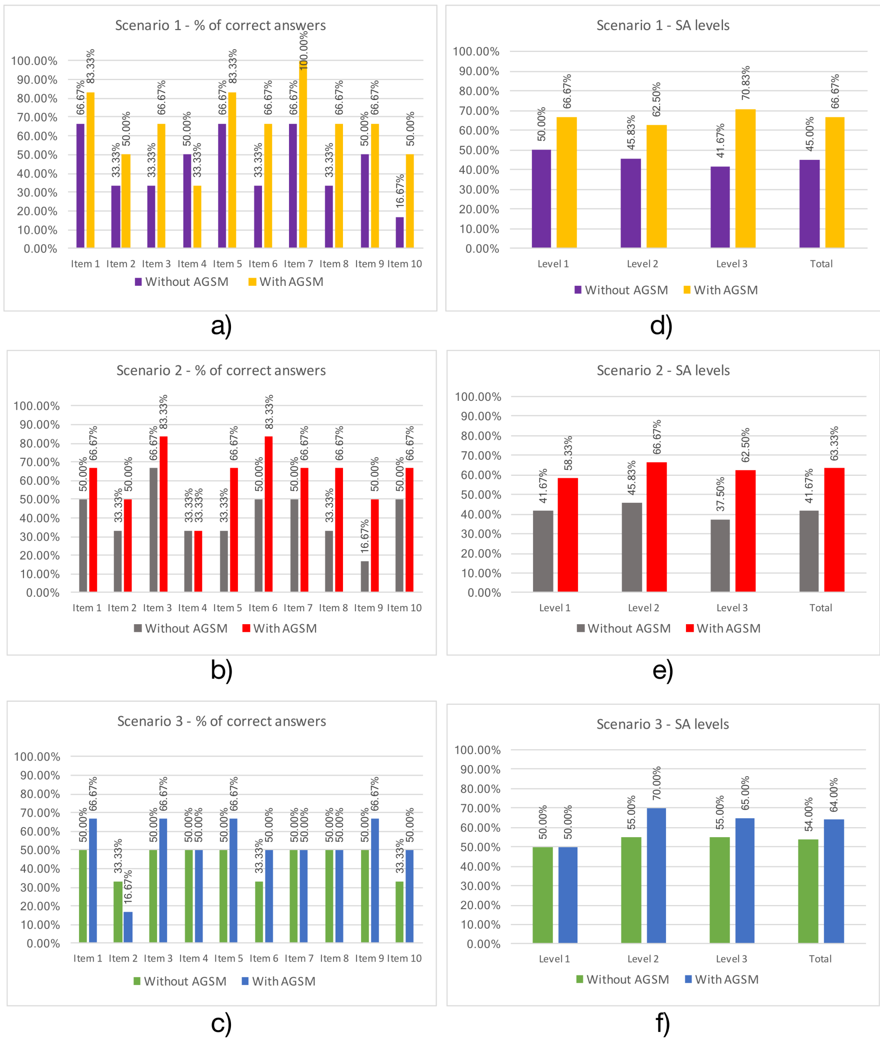


Figure 6.25 Evaluation results of SAGAT applied to GFMS.

lack of experience of the operator, which make a difference in being able to project the situation in the near future for making correct decisions.

Concluding, we can assert that, also in the fleet logistic domain, the AGSM approach can be a valid ally to deal with attentional tunneling and data overload demons, as it effectively supports in switching between goals thus increasing the level of SA even in complicated situations. Anyway, we underline that a proper training of the human operators with respect to the specific application scenario is always needed, as well as the role of the experience is the key for achieving good results. But, also in this cases, the AGSM approach can help unexperienced users to be aware oh what is happening and to be assisted in focusing on the right information.

6.4.3 A DSS for improving SA in the Management of Logistics Port Operations

To further evaluate the support to the SA of the AGSM approach, we evaluate the level of situation awareness gained by the logistics operators by means of a prototypical decision support system (DSS) for the management of the logistics operations in the port container terminal of Salerno, Italy. First, we present the specific scenario of the port container terminal of Salerno. Then we describe the DSS implemented to support the logistics operations related to the handling of the containers in the port. Lastly, we perform an evaluation by means of SAGAT to quantify the level of SA of the users. Furthermore, we propose a numerical simulation to evaluate the impact of the AGSM on the performances of the operations in the port. Therefore, we show that the AGSM approach is useful for supporting the SA and that this, in turn, will allow operators to make better decisions which improve the performance of logistics operations.

Figure 6.26 Aerial view of the port container terminal of Salerno (from Google Maps) with indication of the different yards and their dimensions.

6.4.3.1 Port Container Terminal of Salerno

The port of Salerno is located in the gulf of the Tyrrhenian Sea in the South West of Italy. It is an important port for the international trading of Italy thanks to the favorable geographical position and to a well developed terrestrial connection network. An area of the port is dedicated to the container terminal, as depicted in Fig. 6.26. The choice of using the Salerno container terminal for the case study it is that, due to its small dimensions and the impossibility to enlarge the area due to its position [266], it requires a very efficient management of logistic operations and resources in order to increase the overall performance of the port. The yard capacity is of 12000 TEUs[10], divided in three areas: one for empty containers, one for export containers and the last for import containers (see Fig. 6.26). The containers are moved into and from the yard via reach stackers and shuttles. The load and unload of the containers from the vessels are performed by using 5 cranes. For applying the AGSM approach, we first identify the set of goals for a logistics operator and the related SA requirements by using the GDTA technique. Some of these goals (i.e., those relevant for the evaluation) are reported in Figure 6.27. The overall goal is to achieve an efficient and safe management of the container terminal. To this aim, the manager of the logistics operations have to consider different and competitive operational goals (second level of Fig. 6.27). It should achieve at least a trade-off between increasing the performance of the operations, reducing the overall costs and improving the energy efficiency of the port. Such goals are clearly competitive since, for instance, for increasing the performance, the operators can increase the number of resources involved in the management of a vessel, but this will obviously increase the costs while reducing the energy efficiency.

[10]Twenty-foot equivalent unit, represents the unit of cargo capacity

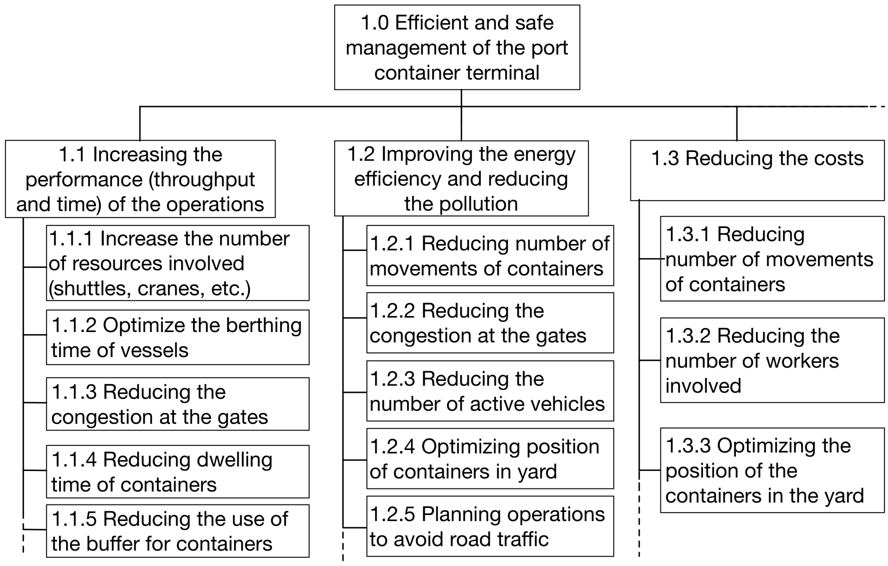


Figure 6.27 Main goals identified by means of the GDTA approach for the case study of the Salerno container terminal.

6.4.3.2 Prototype of the DSS for the Management of Port Logistics Operations

The proposed AGSM approach has been implemented and evaluated by means of a prototypical Decision Support System (DSS) which supports the planning and management of the handling operations of the container in a port terminal. The interface of the DSS is a web application. The architecture and the adopted technologies are the same of the GFMS system described in Section 6.4.2. Fig. 6.28 depicts the main page of the web application, a dashboard summarizing the most important information regarding the port operations and the status of the resources. Using the DSS, the user may control the state of the next arrival or departure of the vessels, as well as the state at the gate-in and gate-out of the port regarding the trucks. Moreover, it is possible to check the state of all the other resources (cranes, reach stackers, shuttles, etc). The information and the operations of the DSS can be organized and visualized according to different views (using different template for the web pages and different kind of widgets).

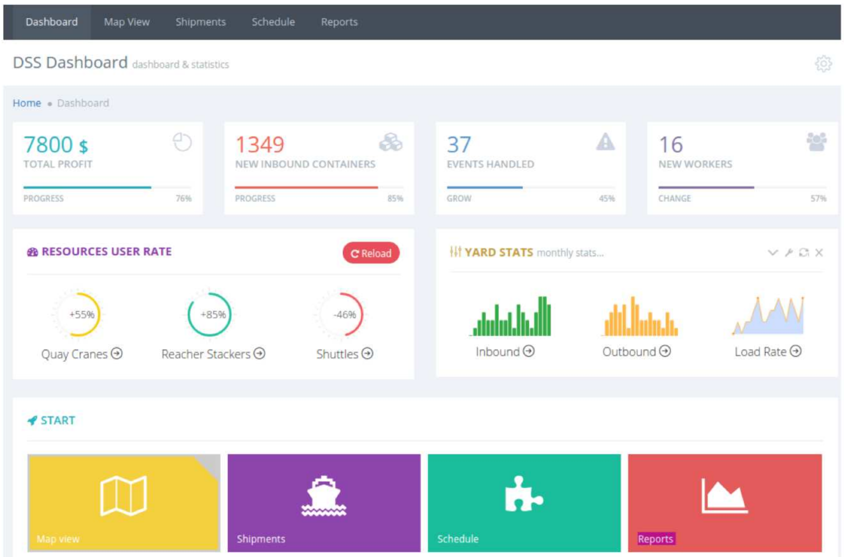


Figure 6.28 Home page of the web application of the DSS

Specifically, to each goal corresponds a different view, and when the user accept the suggestion of changing the active goal, the view of the DSS changes accordingly. For instance, if the goal 1.2 "Improving the energy efficiency and reducing the pollution" is active, the view will show the widgets related to the estimate of the current pollutant emissions, to the number and type of involved resources and the state of the traffic. Fig. 6.29 shows a view of the prototype which contains a plan of operations that should be executed to unload a vessel. The DSS uses different strategies (by means of different optimization algorithms) in order to optimize the operations needed to load/unload a vessel and to optimize the storage in the yard. Specific algorithm can be used according to the current active goal in order to privilege some aspects instead of others (e.g., minimizing the number of movements in order to improve the energy efficiency and reduce the pollution). Specific events and alarms may arise while the user is interacting with the system. For instance, an unexpected breakdown of a crane or a delay in the arrival time of a vessel. Such kind of data-driven events are notified via a notification tab and pop-ups. The user

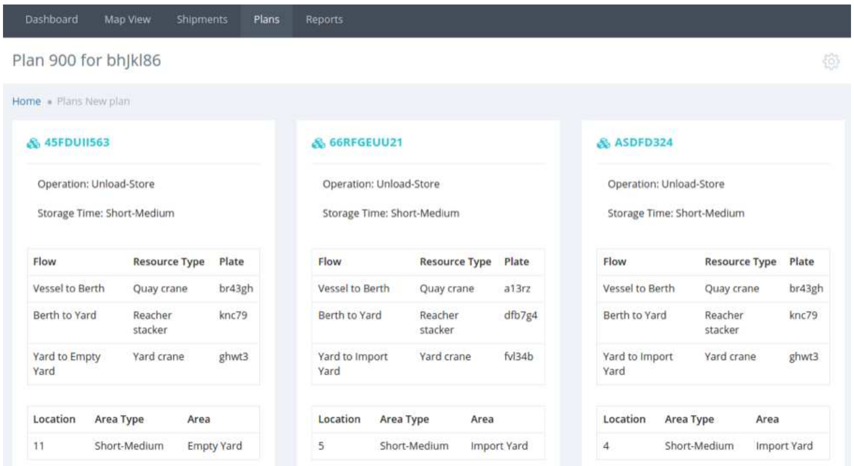


Figure 6.29 Plan of operations to unload a vessel generated by the DSS

is asked to manage such events, by using the interface shown in Figure 6.30, in which the user can select the best strategy to solve an issue. According to the AGSM approach, when a goal becomes more desirable than the active one, a suggestion to change the active goal is sent to the user. In the prototype, this suggestion is shown via a notification tab. The user may select the suggested goal or another one by using the interface of 6.30. This will cause a change of the view of the DSS.

6.4.3.3 SAGAT Evaluation

The first evaluation of the DSS is performed by using SAGAT to measure the level of SA gained by the operators thanks to the AGSM approach.

6.4.3.3.1 Data and Method. Two different versions of the DSS are used for the evaluation:

- DSS-A) the DSS does not implement the AGSM appraoch, thus it does not suggest to change the active goal;
- DSS-B) the DSS implements the AGSM approach and it suggests to change the active goal.

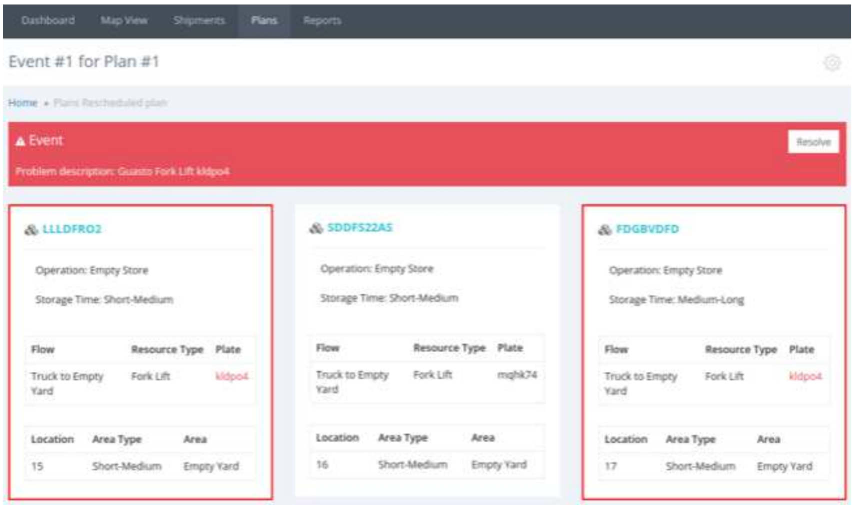


Figure 6.30 View for the selection of the strategy to solve a specific problem and for the selection of the goal to activate.

The sample consists of 15 users with experience in the port logistics operations. First, the users have been trained in using the system. Then, each of them had the task of coordinating and managing all the operations for the load/unload of the vessels in three different simulation scenarios:

- Scenario 1) only one vessel to load/unload;
- Scenario 2) simultaneously unload a vessel and load another one;
- Scenario 3) load and unload two vessels with transshipments operations.

These scenarios have been defined in the context of the Research&-Development project (funded by the Italian Ministry of Instruction, University and Research) Mar.Te[11]. Each scenario is executed twice with the two different versions of the system, DSS-A and DSS-B. The questionnaire contains 30 queries: 15 to assess the SA at the perception level, 10 for the comprehension level and

[11]http://mar-te.com and http://www.corisa.it/project/martem/

5 for the projection level. At the perception level, the queries aim at identifying if the user has perceived single pieces of information, asking for instance: number of shuttles involved in the unload of the vessel; number of involved cranes; the percentage of occupancy of the yard (even roughly); average time for a truck to enter from the gate, and so on. At the comprehension level, the queries aim at identifying if the user has correctly understood the meaning of the information perceived at level 1 according to his/her goal, asking for instance: 'are there other free reach stackers that can be used to unload the vessel?'; 'is the vessel waiting too much before the unloading operations starts?'; 'is it possible to perform the transshipments of some containers without using the buffer?', and so on. At the projection level, the queries aim at identifying if the user is able to predict the future state of the situation, asking for instance: 'considering the current number of resources involved, is it possible to respect the arrival time of the next planned arrival of vessel?'; 'By stopping two reach stackers from the unloading operations, are you able to guarantee that the vessel will be unloaded in an hour?', and so on.

Even in this case, a subset of queries is randomly selected for each simulation and each user and submitted at random time instant in which the simulation is frozen. The queries are selected in order to ensure that we obtain at least 30 answers for each query and a maximum of 45 answers for each of the two modalities of the DSS.

6.4.3.3.2 Results and Discussion. Figures 6.31.a-c show, for each scenario, the average of correct answers grouped together according to the level of Situation Awareness to which the questions refer to. For each level of SA, the graph compares the percentage of correct answers with respect to the two versions of the DSS, DSS-A and DSS-B. All the three levels of SA are improved thanks to the adaptive goal selection approach, with an average percentage increase of 14.59% of correct answers. Specifically, in the first two scenarios we obtain a clear improvement for all the three levels of SA, while in the third scenario (Figure 6.31.(c))

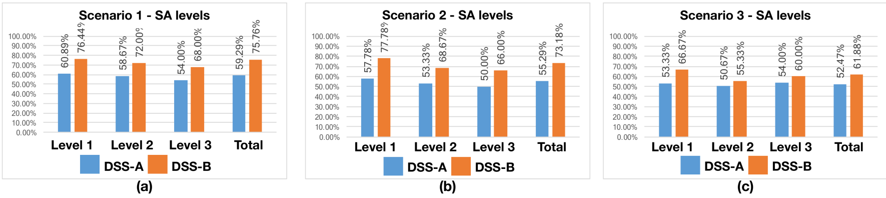


Figure 6.31 Evaluation results. Each graph refers to a scenario and shows the percentage of correct answers.

we observe just a slight increase in correct answers at SA level 2 (Comprehension) and SA level 3 (Projection). The number of operations to manage in the last scenario is huge; moreover, the operations for transshipping hinder the task of the user as it requires a big effort in order to simultaneously manage the load and unload of two vessels. Since the users are not very familiar with this task, they are not able to adopt the correct mental model for understanding the current situation and the required experience to predict and plan future operations. This means that a fundamental role is still played by the user experience confirming that SA training is crucial to the success of any system for SA[9]. In conclusion, also in this domain and for the task of handling container in a port terminal, the AGSM approach provides good results in the improvement of the SA gained by logistic operators.

6.4.3.4 Performance Evaluation with Numerical Simulation

The second evaluation of the DSS aims at verifying if the AGSM approach is able to improve the overall performance of the operations in the container terminal of Salerno.

6.4.3.4.1 Data and Method. The evaluation is performed by simulating the operations of the Salerno Container Terminal. The simulation is realized by means of Arena® Simulation Software[12] in which we have implemented the simulation model of a container terminal proposed in [266]. The discrete event-based

[12] Arena Simulation Software www.arenasimulation.com

simulation we have implemented allows modeling the workload of the terminal (e.g., arrivals and departures of vessels and trucks, number of containers, etc.) with different kind of probability distributions and then to evaluate different Key Performance Indicators (KPIs). For this evaluation, we consider some KPIs that are useful to evaluate the efficiency of the operations, specifically by considering:

- KPI_{V-lut}: Vessel load/unload time
- KPI_{Cm}: Total number of movements of the containers in load/unload operations
- KPI_{cr}: Number of cranes used for a single vessel
- KPI_{S-wt}: Shuttle waiting time
- KPI_{SV-lt}: Loading time of full containers from Shuttle to Vessel
- KPI_{BV-lt}: Loading time of full containers from Berth to Vessel
- KPI_{BY-tt}: Transfer time from berth to yard of a container

We compare the average values of the KPIs along two simulations: one in which we do not consider the adaptive goal selection approach (SIM-A) and the other one in which we consider the goal selection approach (SIM-B). Specifically, this means that in SIM-A we perform a simulation to obtain the baseline scenario of the experiment, using the workload and the parameters provided in [266] for the port of Salerno. Such workload characterization and parameters have been identified by the authors of [266] by gathering real data that refers to the period between January 2003 - July 2005 (more than 1.000 vessels) and by an integrative survey carried out during six months in 2005 to gather data about the movements of containers on the berth. In SIM-B we consider that a hypothetic operator, responsible for the optimization of the logistics operations, always accepts to focus on the goal suggested by the Adaptive Goal Selection approach. When a specific goal becomes active, we modify the parameters of the simulation. In particular, we focus on the three goals of the GDTA in Fig. 6.27. When goal 1.1 is active, the aim is to increase the performance. In this case, we use the maximum number of available resources (cranes, shuttles, etc.) to minimize the operations time and maximize the throughput. When goal 1.2 is active, the aim is to increase the energy efficiency. In this case, we try to reduce the number of resources to use and we prefer the most efficient resources. When goal 1.3 is active, the aim is to reduce the costs. In this case, we want to find a trade-off between time and the number of workers and resources.

6.4.3.4.2 Simulation of the Salerno Container Terminal in ARENA. The simulation of the Salerno Container Terminal is implemented in Arena® as a discrete event system. This system can be represented as a graph in which the nodes are the significant events, the edges are the activities (e.g., movements of a container) and a path is a sequence of activities. In such a graph, the containers move following different paths according to their destination. The structure of the discrete event simulation model allows to easily compute the key performance indicators since they are functions of the model variables. The simulation is based on the model of the Salerno Container Terminal proposed in [266], where the authors define all the events, resources and activities involved when loading and unloading the vessels. In Fig. 6.32 we report a conceptual view of the simulation model divided into 6 sub-models:

- *Arrival*: arrivals of vessel to load/unload
- *Unloading*: unloading of the containers from the vessel to the berth by using a crane.
- *Transshipment*: unload of a container form a vessel to the berth and from the berth to another vessel.

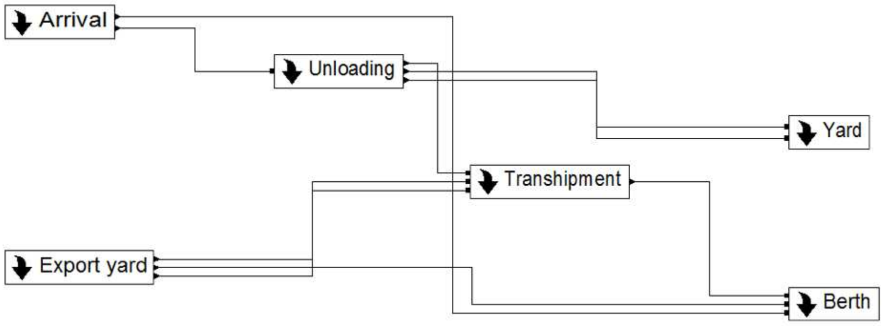


Figure 6.32 Conceptual view of the simulation model for the Salerno Container Terminal.

- *Yard*: the unloaded containers that must not be loaded on another vessel are moved into the yards by means of reach stackers, shuttles and forklifts.

- *Export Yard*: flow of the containers that reach the terminal through the gate.

- *Berth*: movements of the containers on the berth, which is used during the load/unload of the vessel.

Fig. 6.33 is an extract of the simulation model implemented in Arena which refers to the Yard sub-model.

The pink boxes represent the *stations* which means that the containers reach this location and wait to be processed (for instance, to simulate the berth or the yard). The blue boxes represent different events and activities. For instance, a blue box can represent a resource request (e.g., Shuttle Request) in which the container waits until a requested resource is free. They can also represent the movements of the containers (e.g., transport from shuttle station to import yard) in which it is possible to simulate the time needed to perform a movement. The rhombus represent the decision branches (e.g., in Fig. 6.33, the "Rail"indicates if the container should be transported outside the yard by train). The decision of which branch to consider next, is made according to a probability.

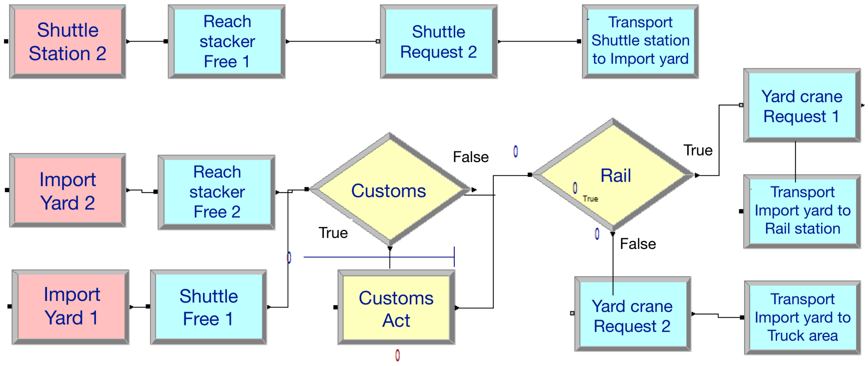


Figure 6.33 Details of the Yard sub-model implemented in Arena

6.4.3.4.3 Results and Discussion. The simulation model has been calibrated with the parameters proposed in [266]. We performed the experiments by using the two modalities of the simulation model (SIM-A and SIM-B), and by executing 25 simulations [266] for each modality. We considered the average values of the KPIs above described. Table 6.4 summarizes the results of the evaluation. Comparing the two modalities, it is clear that the adaptive goal selection approach provides (averagely) an improvement of the performance. Specifically, when considering all the time-based KPIs, we obtain an average percentage improvement of -13.74% for the *mean* values, -7.6% for the *min* values and -9.3% for the *max* values. Moreover, we observe a significant improvement both in the load/unload time of vessels (-18.18% on the *mean* values) and in the loading time from berth to the vessels (-34.22% on the *mean* values), demonstrating that the AGSM approach can be useful to improve the performance of such operations.

6.5 Summary

When dealing with complex systems and dynamic environments, the quantity of information to perceive and understand quickly outpaces the cognitive capability of humans, due to the well-known

Table 6.4 Average values of the KPIs. In bold the best results.

	SIM-A			SIM-B		
KPI	*min*	*mean*	*max*	*min*	*mean*	*max*
KPI_{V-lut} (hour)	1.1	13.2	**56.2**	**0.8**	**10.8**	59.1
KPI_{Cm} (cont./vessel)	**4**	295.8	1579.2	4.1	**276.2**	**1522.3**
KPI_{cr} (crane/vessel)	1.0	1.6	**3.0**	1.0	**1.5**	3.2
KPI_{S-wt} (minutes)	0.21	2.54	**9.87**	**0.19**	**2.21**	11.89
KPI_{SV-lt} (minutes)	0.43	**1.43**	3.24	**0.39**	1.44	**3.11**
KPI_{BV-lt} (minutes)	0.67	1.87	3.45	**0.64**	**1.23**	**3.34**
KPI_{BY-tt} (minutes)	**0.98**	1.49	4.2	1.1	**1.43**	**1.47**

data overload problem. A way to deal with such complexity is to process information according to a specific goal, leveraging on the so-called goal-driven information processing. Having a goal to achieve, the human is able to filter only the information that are relevant, thus reducing the amount of data that should be processed. Unfortunately, this can cause the problem of attentional tunneling, as the human may lose the global view on the overall system, thus not perceiving critical information about the state of the environment since they are not directly related with the current goal. To avoid this problem, specific cues of the system or of the environment should capture the attention of the user on such important information, leading the human to process information with a data-driven approach (i.e., without a particular goal to achieve). Unfortunately, this approach is less effective than the goal-driven, and it can bring back to the problem of information overload. How to deal with such interdependencies between attentional tunneling and data overload demons? The answer is in finding the right trade-off between goal-driven and data-driven information processing, supporting the humans in switch coherently between these two modalities and by using wisely the salience of information to recall their attention.

Accordingly, in this chapter we propose a computational approach for situation management, namely Adaptive Goal-driven Situation Management, that supports this trade-off by means of a process for selecting the most desirable goal on which the human operator should focus the attention. A formal model of goals and situations has been defined in order to sustain the process of goal selection and the support to the execution of the active goal, fostering a shared understanding and a collaboration between software agents and human operators. AGSM adopts a reinforcement learning algorithm in order to learn from users/operators' feedback the ability to suggest alternate goals.

The evaluations conducted by means of three prototypical systems in different domains, demonstrate that the approach is capable of increasing the SA of the human operators of such systems, thus effectively dealing with the aforementioned demons of attentional tunneling, data overload and misplaced salience demons.

Chapter 7

Improving Situation Awareness with Granular Computing

> *"All things are subject to interpretation.*
> *Whichever interpretation prevails at a given time is a function of power and not truth."*
>
> — Friedrich Nietzsche

In chapter 3 we analyzed the main principles and techniques of Granular Computing (GrC) and we paved the way for an integration of granular computing in situation awareness, pointing out that GrC can be effectively used to support all the phases of SA. In this chapter we propose a set-theoretic framework for modeling situations with GrC, supporting situation perception, comprehension and projection.

The definition of computational models of situations is crucial for the realization of SA systems. As we evidenced in chapter 2, many models for representing situations have been proposed so far, based on the most different approaches, ranging from formal logic to machine learning. The way by which a situation is represented in an SA system directly influences all the other processes and capabilities of the system The choice of adopting a model instead of another one depends on the specific application and users' goals. Basically, when choosing a model for representing situations, a trade-off should be found between the expressiveness, the formalism, the capability to directly 'see' and understand the model, which are typical characteristics of expert-based approaches (e.g., formal logic, ontologies) and the capability of dealing with uncertainty, extracting patterns from raw data, the flexibility and scalability of learning-based approaches (e.g., neural networks, Hidden Markov Models). Given the characteristics and principles of the Granular Computing (GrC) which provides human-inspired problem-solving approaches together with computational techniques supporting them, it can be useful to model situations with granular structures. In such a way, one can benefit from the expressiveness of the visual representation of granular structures by means of lattices of partitions, and the capability of dealing with uncertainty, representing de facto a good trade-off between expert-based and machine learning approaches.

Parts of this chapter have been previously published in:

- V. Loia, G. D'Aniello, A. Gaeta, F. Orciuoli (2016). Enforcing situation awareness with granular computing: a systematic overview and new perspectives. Granular Computing, 1(2), 127-143.
- G. D'Aniello, A. Gaeta, V. Loia, F. Orciuoli (2017). A granular computing framework for approximate reasoning in situation awareness. Granular Computing, 2(3), 141-158.
- G. D'Aniello, A. Gaeta, V. Loia, F. Orciuoli (2017) A model based on rough sets for situation comprehension and projection. 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management, CogSIMA 2017, (***Best Paper Award***)

7.1 Situations as Granular Structures

The proposed model of situations is based on a set-theoretic framework for GrC. In this model, we use the concept of information granule as a way to improve the perception of the elements of an environment by grouping together such elements according to a criterion of similarity, proximity, indistinguishability, or any other requirements imposed by the specific SA application. A situation is represented as a *granular structure*, a structure in which different information granules are organized by means of specific relations. Such structures ease and improve the perception of the information coming from the environment as they reduce the complexity and the amount of data by representing many elements with a unique granule. Moreover, the structures can be directly visualized and manipulated by the human operators, thus supporting the direct perception of information at Level 1 and Level 2 SA, thus implementing one of the design principles indicated by Endsley and Jones to solve SA demons [15]. The situation model also supports Level 3 SA, since different granular structures can be compared in order to quantify the differences among them. In such a way, it is possible to understand how the current situation is evolving over the time by comparing and measuring the dissimilarity (or the distance) between the granular structures representing the current situation and the one representing a plausible projection. Furthermore, the proposed situation model can be exploited to support the reasoning on situations by means of a *conformity analysis* between a recognized situation and an expected one. By exploiting this capability, it is possible to automatically alert human operators when two situations are not compliant. Another powerful capability of modeling situations with granular structures is their support in dealing with uncertainty. Uncertainty is everywhere in real-world scenarios, especially when dealing with data gathered by physical sensors. Granular structures give the possibility of dealing with different kinds of uncertainty by means of different theoretical framework (sets, fuzzy sets, rough sets, clustering, etc.).

7.1.1 Granulation Approach for creating Granular Structures

As outlined in chapter 3, the process for creating granular structures starting from raw data is called granulation. Figure 7.1 depicts the process of granulation for creating granular structures that represent situations by abstracting raw data according to the SA requirements identified in the first phase of the approach. According to the adaptive goal-driven situation management approach proposed in chapter 6 and depicted in Figure 6.1, in the initial steps of the definition of an SA system it is needed to analyze the requirements (from an SA perspective) by leveraging on the two complementary approaches of structured thinking, i.e., top-down (goal-driven) and bottom-up (data-driven). Specifically, as depicted in Figure 7.1, in the goal analysis we usually adopt the Goal-directed Task Analysis (GDTA) [15]. The results of this phase of analysis are a set of SA requirements, a hierarchy of goals, a set of tasks and other information requirements that will drive the granulation process to identify the granular structures for representing situations. The granulation process starts from raw data gathered by sensors (see Figure 7.1). Such data can be grouped to create *type 1* granules. The criteria by which such granules should be created come from the information requirements (of level 1 SA) elicited in the GDTA process. Typically, at this level, common issues relate to object recognition, feature reduction and outlier detection, and the granules should be created to address such issues. Different techniques and approaches can be exploited to create level 1 granules, like the ones described in Section 3.2.1.1. Different granules of type 1 can be organized in a hierarchical structure, according to the information requirements. The number of granules and levels to be created depends on the number of elements to be perceived from the environment, on the number of their features, on the complexity of the object and so on. Furthermore, also the type of formalism used to describe the granules and the kind of relation used (e.g., similarity, distance, functionality) have an impact on the resulting granular structure

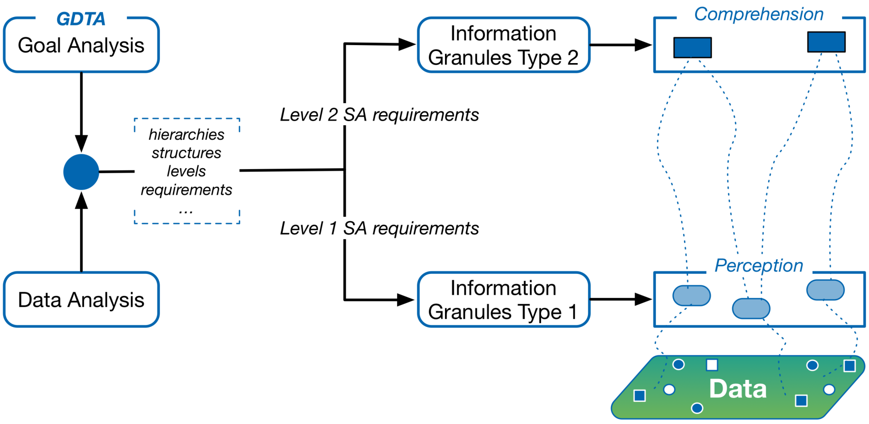


Figure 7.1 Overall approach for representing situations with granular structure

at this level. Type 1 granules (and granular structures) can be further abstracted and organized to create *type 2* granules. Such granules are created according to the information requirements of comprehension level (level 2 SA) defined in the first phase of the approach, both by means of the GDTA and by means of bottom-up approaches. Coarsening (or abstraction) relationships can be used to create type 2 granules from type 1 granules, with a bottom-up approach. As also specified in Section 3.1.3, we can also use a top-down approach (or a middle-out), thus starting from level 2 requirements of the GDTA and defining the type 2 granules, and then using some refinement relationship to create the type 1 granules.

A measure for taking into account the degree of uncertainty or imprecision that characterizes a granule (due to the fact that it clumps together many elements) is the *information granularity* IG. Information granularity gives a measure of how much information is granulated (e.g., grouped together) in a structure. A finer granulation corresponds to a low value of IG. A too high value for IG may represent a too coarse or imprecise granulation, while a too small value may represent a very precise representation of the elements of the environment, and consequently, it is not so useful for SA applications requiring information fusion since we are simply considering individual elements. The appropriate level of information granulation depends both on the SA information requirements and on the capabilities of the human operators such as their attention level, memory, task load, etc. Notice that information granularity is also a human-understandable measure of how much information is contained into a granular structure, as the human operator may evaluate the degree of abstraction of situations by comparing their values of IG.

7.1.2 Evolvable Situations

Type 1 and Type 2 granular structures allow implementing the SA requirements of level 1 and level 2, since they give us a snapshot of a situation at a given time. The SA requirements of level 3 (i.e., projection) can be accommodated by means of the so-called *evolvable granules* [267, 268]. The definition of situation awareness given by Endsley focuses on the coupling between time and space dimensions. Indeed, situations change over time and are related to a given space of the environment. Supporting SA by means of granular structures thus requires considering the coupling between time and space in the granulation processes. Following the work of Leite et al. [269], we first consider time granulation and then space granulation. This implies that, when considering data coming from sensor networks, the sensor readings are analyzed in a given fixed time window. We granulate the time by fixing a time window γ and then we perform space granulation of the slice of data in this time window. The result of the space granulation is a granular structure of two levels of granularity ϵ and ϵ^I corresponding to the two levels of SA requirements (perception and comprehension). When selecting a different time window γ^I, a different granular structure is obtained, wherein some granules can be merged, split or new granules can be created. This is the behavior of evolvable granules [268].

Figure 7.2 depicts a simple example of the process of time and space granulation. The figure shows three time windows of width=γ; in the first time window, the sensors register the position

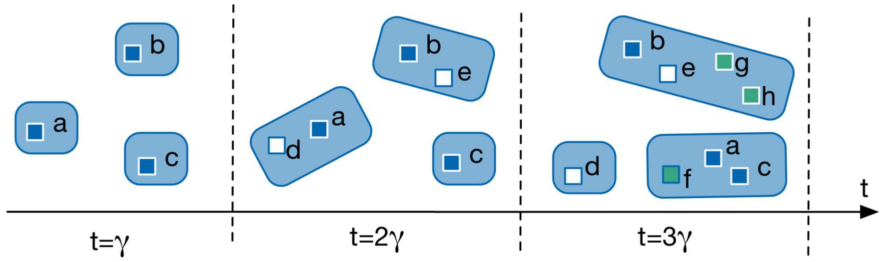


Figure 7.2 Granulation along time and space dimensions

in the space of the objects a, b and c; in the second of d, e and in the third f, g, h. Supposing we use a criterion of distance for creating the granules. The figure shows the granular structures (depicted with a blue box encapsulating the objects) in each time window, thus representing an evolvable granular structure during the time. Each granule reports the position of the objects in the space.

In the first time window, we have three granules (at level 1). In the second time window, the objects d and e can be merged with the existing granules (due to their proximity with such granule), thus creating the granules $\{a, d\}$ and $\{b, e\}$. And this process of granulation evolves during the subsequent time windows and, as output, we obtain evolvable granular structures in which the granules can be merged, split removed or added. Reasoning on such evolution of granular structure can support the SA requirements of projection level (level 3).

Figure 7.3 depicts a graph that shows the evolution of the granular structures of Figure 7.2, combining time and space granulation. The tree in the middle of the figure reports time granulation. It is a lattice of partitions of indistinguishable objects in time slices of different width. Thus γ, 2γ and 3γ are the dimensions of the time slice, thus considering a time granulation with different sizes of the time windows. When considering a time slice of width γ (top of the figure) a, b, and c are indistinguishable with respect to time. Indeed, along with the time dimension, they are a single granule $\{a, b, c\}$. The same is for granule $\{d, e\}$, as well as for $\{f, g, h\}$. If we consider a larger time slice, e.g. 2γ (depicted in the middle level of the tree structure), we have coarser gran-

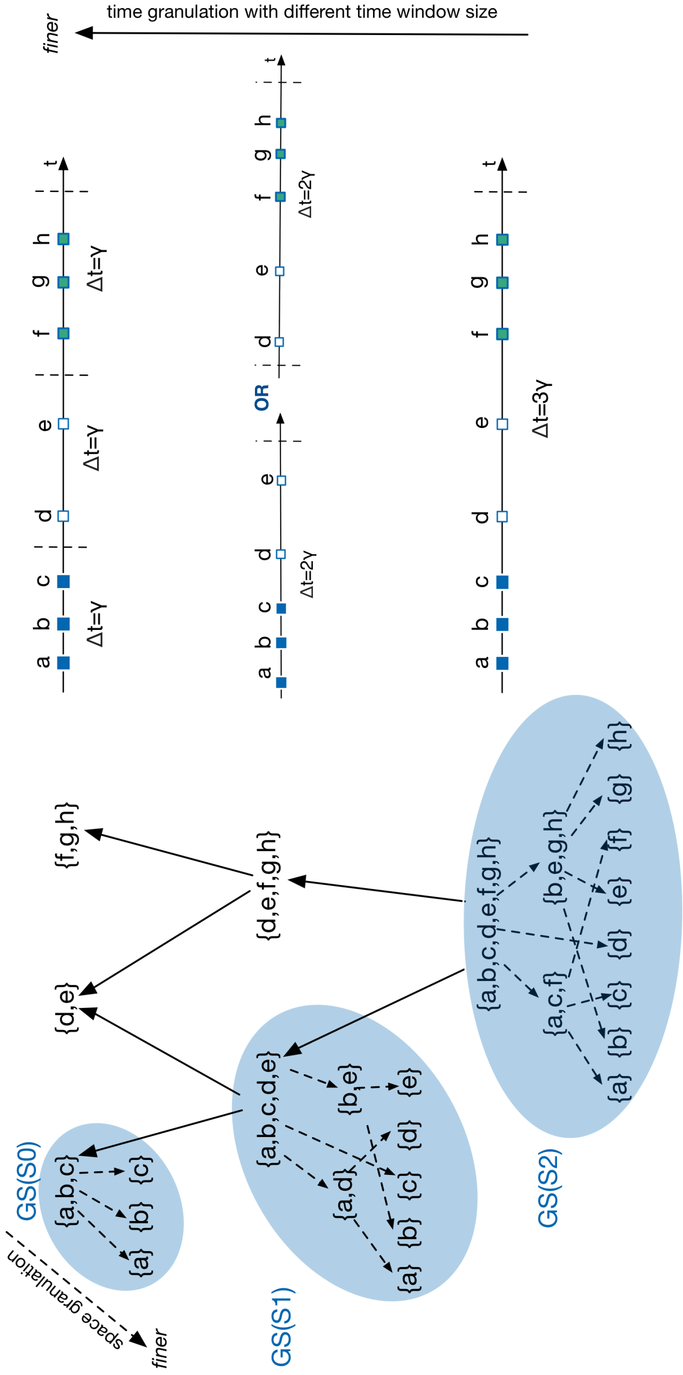


Figure 7.3 Evolution of granular structures

ules of indistinguishable objects, e.g. $\{a, b, c, d, e\}$ or $\{d, e, f, g, h\}$ (the different granules are obtained by considering a different start of the time window, as depicted on the right of the figure). The coarsest granulation with respect to the time is for a time windows of 3γ, where we have only one granule containing all the objects. The grey ellipses show the results of space granulation for the case reported in Figure 7.2. Each ellipse gives a snapshot of a situation at a specific time-slot represented by a granular structure.

Intuitively, to accommodate in a correct way SA L3 requirements, SA operators have to reason on the transitions $GS(S_0) \rightarrow GS(S_1) \rightarrow GS(S_2)$. In such case, it is also useful to evaluate how much a projected situation (i.e., a possible evolution of the current situation) differs from the recognized one, in order to take a decision according to the possible evolutions. To this purpose, the concept of distance between granular structures is used to evaluate the dissimilarity between two granular structures representing consequential snapshots of a situation.

7.2 A Theoretical Model for representing Situations with Neighborhood Systems

The concepts of granular structures and evolving granules can be exploited to represent situations and their evolutions over the time as intuitively discussed in the previous section. More formally, we define a theoretical framework based on the concepts of neighborhood systems (introduced by Yao in [270]) for representing the situations by means of the granular structures. Given an element x of the universe U and a distance function $D : U \times U \rightarrow \mathbb{R}^+$, for each $d \in \mathbb{R}^+$ we define the neighborhood of x as:

$$n_d(x) = \{y | D(x, y) \leq d, y \in U\} \tag{7.1}$$

The set $n_d(x)$ of Eq. 7.1 is a type 1 granule. A cluster (or a set) containing the element x can be represented by $n_d(x)$. Equation

(7.1) is generic enough to support also other types of granulation besides spatial proximity. Specifically, considering D as similarity function, $D : U \times U \rightarrow [0, 1]$, Eq. 7.1 defines a granule of similar elements. When considering D as an equivalence relation, Eq. 7.1 denotes an equivalence class. Again, considering D as a fuzzy binary relationship, $n_d(x)$ will represent the neighborhood as a fuzzy set.

From Eq. 7.1, high order granules and granular structures may be constructed. Let us consider a neighborhood system NS of x as a non-empty family of neighborhoods:

$$NS(x) = \{n_d(x) | d \in \mathbb{R}^+\} \tag{7.2}$$

Neighborhood systems can be used to create multi-layered granulations. Specifically, considering a nested neighborhood system:

$$NS(x) = \{n_1(x), n_2(x), ..., n_j(x)\} \tag{7.3}$$

with $n_1(x) \subset n_2(x) \subset ... \subset n_j(x)$, induces a hierarchy of neighborhoods (granules) of x, which allows to define refinement and coarsening relationships on granules $n_1(x) \prec n_2(x) \prec ... \prec n_j(x)$. The union of neighborhood systems for all the elements of an universe defines a granular structure:

$$GS = \cup_{i=1}^{|U|} NS(x_i) \tag{7.4}$$

If $NS(x_i)$ is a hierarchy (i.e., it is defined as a nested neighbhorood system as in Eq 7.3), GS is a hierarchical granular structure.

For a granular structure GS we define the information granularity as:

$$IG(GS) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{|NS(x_i)|}{|U|} \tag{7.5}$$

whose extremes (corresponding to the finest and coarsest granularity) are $\frac{1}{|U|} \leq IG \leq 1$.

As we evidenced in the previous section, the distance between two granular structures (representing two situations or a situation and its projection) is useful for the human operator to understand how much the situations differ (improving the SA comprehension and the projection phases). To define the distance between two granular structures, we borrow the definition used for rough sets [271] and fuzzy sets [272]. Given two granular structures GS_1 and GS_2 we define their distance as follows:

$$D(GS_1, GS_2) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{|NS_1(x_i) \triangle NS_2(x_i)|}{|U|} \tag{7.6}$$

where $|\cdot|$ is a cardinality, and

$$|NS_1(x_i) \triangle NS_2(x_i)| = |NS_1(x_i) \cup NS_2(x_i)| - |NS_1(x_i) \cap NS_2(x_i)| \tag{7.7}$$

is the cardinality of a symmetric difference between the neighbourhood systems. Eq. 7.6 is a distance measure because the operation $\triangle$ removes the elements that are common between two sets and, thus, can be considered as a sort of dissimilarity. In Eq (7.6), the accumulated dissimilarity between all the granules of two granular structures is considered, thus representing a measure of their distance.

Since (7.6) is a measure of dissimilarity between two granular structures, a measure of similarity can be defined as:

$$S(GS_1, GS_2) = 1 - D(GS_1, GS_2) \tag{7.8}$$

7.2.1 Example of modeling Situations with Neighborhood Systems

The following example (based on an operational scenario extracted from [273]) shows how the granular structures are created by using the proposed set-theoretical model. A flight air traffic controller monitors flight paths in order to assess rare events or unusual situations. We suppose an unusual situation that should be identified. Specifically, a splitting maneuver of an aircraft can be a dangerous situation that happens when one aircraft, staying close in a group, suddenly moves away from a predefined trajectory. This

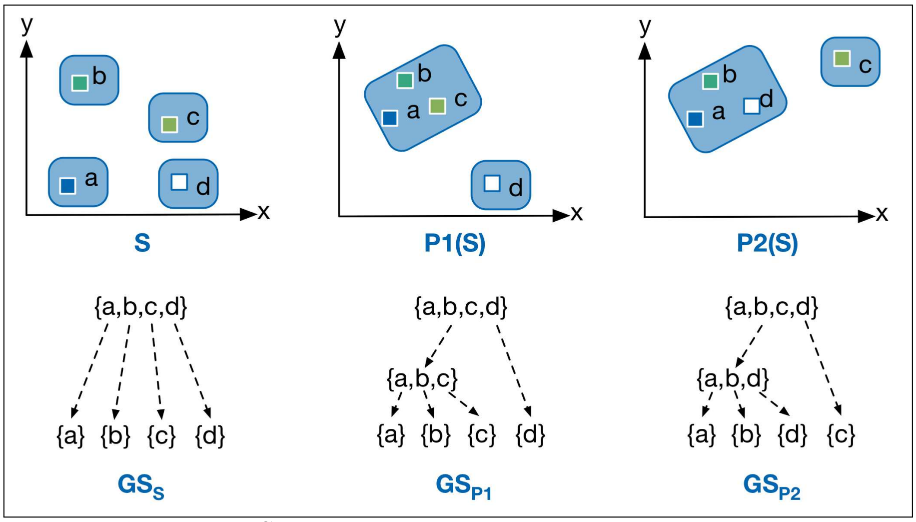


Figure 7.4 Situations and granular structures - Example

can represent a problem with the aircraft (failure), a hijacking, an attack, etc.

Let us suppose $U = \{a, b, c, d\}$ is the set of all aircrafts. In the situation S at time $t = t_0$, each aircraft is distant from the others (separated objects). Let us suppose that from S the human operator expects two probable projections, let us call $P1(S)$ and $P2(S)$, where three objects group together. The example is graphically shown in Figure 7.4.

Using Eq. (7.2), in S the following neighborhood systems exist:

$$
\begin{aligned}
NS(a) &= \{\{a\}\} \\
NS(b) &= \{\{b\}\} \\
NS(c) &= \{\{c\}\} \\
NS(d) &= \{\{d\}\}
\end{aligned}
$$

Accordingly, using Eq. (7.4) the granular structure GS_S is the union of four singletons corresponding the the four objects of the universe.

For $P1(S)$ we have the following neighborhood systems:

$$\begin{aligned} NS(a) &= \{\{a\},\{a,b,c\}\} \\ NS(b) &= \{\{b\},\{a,b,c\}\} \\ NS(c) &= \{\{c\},\{a,b,c\}\} \\ NS(d) &= \{d\} \end{aligned}$$

In this projection, $NS(a)$, $NS(b)$ and $NS(c)$ are nested systems and induce the hierarchy $GS_{P1(S)}$ in Figure 7.4 with the creation of a coarse granule $\{a,b,c\}$ reporting information on a group of aircrafts.

For $P2(S)$ we have the following neighborhood systems:

$$\begin{aligned} NS(a) &= \{\{a\},\{a,b,d\}\} \\ NS(b) &= \{\{b\},\{a,b,d\}\} \\ NS(c) &= \{c\} \\ NS(d) &= \{\{d\},\{a,b,d\}\} \end{aligned}$$

and the granular structure for this second projection $GS_{P2(S)}$ seems to be, in some way, similar to $GS_{P1(S)}$ for what concern the number of the aggregated objects (see Figure 7.4).

To quantify the difference between the granular structures, we calculate the information granularity with Eq. 7.5:

$$\begin{aligned} IG(GS_S) &= \frac{1}{4}[\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}] = \frac{1}{4} \\ IG(GS_{P1(S)}) &= \frac{1}{4}[\frac{2}{4}+\frac{2}{4}+\frac{2}{4}+\frac{1}{4}] = \frac{7}{16} \\ IG(GS_{P2(S)}) &= \frac{1}{4}[\frac{2}{4}+\frac{2}{4}+\frac{1}{4}+\frac{2}{4}] = \frac{7}{16} \end{aligned}$$

and the distance between these structures with Eq. 7.6:

$$
\begin{aligned}
D(GS_S, GS_{P1(S)}) &= \frac{1}{4}[\frac{2-1}{4} + \frac{2-1}{4} + \frac{2-1}{4} + 0] = \frac{3}{16} \\
D(GS_{P1(S)}, GS_{P2(S)}) &= \frac{1}{4}[\frac{3-1}{4} + \frac{3-1}{4} + \frac{2-1}{4} + \frac{2-1}{4}] = \frac{6}{16} \\
D(GS_S, GS_{P2(S)}) &= \frac{1}{4}[\frac{2-1}{4} + \frac{2-1}{4} + 0 + \frac{2-1}{4}] = \frac{3}{16}
\end{aligned}
$$

Now, let us analyze the measures we have calculated and the granular structures depicted in Figure 7.4, to point out the capabilities of the proposed theoretical situation model to support all the three levels of SA.

- **Supporting perception:** the definition of a granule as a neighborhood of a given element x with Eq. 7.1 fits with the SA level 1 requirements. Specifically, in the proposed example, the requirement is to perceive the elements that are spatially close, and so the function $D(x, y)$ is a distance function. The same concept of neighborhood can be applied to other requirements. When the requirement is to perceive similar objects, the function $D(x, y)$ is a similarity function; when the requirements demand to perceive indistinguishable objects, the function $D(x, y)$ is an equivalence relation. Thus, the added value in defining type 1 granules with the proposed model relies on the flexibility of using different granulation criteria according to the SA level 1 requirements.

- **Supporting comprehension:** the SA level 2 requirements usually regard the understanding of some kind of relation among the objects perceived at level 1. In the example, we should understand if the objects (aircrafts) are spatially close together to comprehend the situation. In the proposed model, this is supported by the creation of neighborhood systems NS (Eq. (7.2)). In fact, the NS can be further aggregated into a multi-layered hierarchical structure GS (Eq.

(7.4)) which is an approximated representation of the elements at a particular time. A measure of the uncertainty associated with such representation is given by the information granularity IG (Eq. (7.5)). The elements perceived at level 1 are organized in a granular structure according to the level 2 requirements, thus improving the comprehension of a situation, since it also represents the relations among the elements graphically. Moreover, the information granularity and the distance between granular structure gives important information for the human operator. The first gives an indication about the uncertainty associated with the granular structure (and in the case in which the IG is high, this can drive the search for gathering further information and for performing a deeper analysis). The distance between granular structures provides a measure to understand how much two different situations differ, and thus the human operator may rapidly understand if the situation is changed or it remains the same as time passes.

- **Supporting projection:** the information granularity IG and the distance D between granular structures support the projection level by providing the operators with early indications on how a projected situation differs with respect to the recognized one at level 2. In the proposed example, the information granularity of the recognized situation is $IG(GS) = \frac{1}{|U|}$ which is the finest granulation. This means that the operator has a complete and precise information about the position of the four objects. This value is different from the information granularity of the two projections $P1(S)$ and $P2(S)$. This means that both projections bring new and additional information which is represented by the creation of a coarser granule. However, the information granularity of the two projections is the same, i.e. $IG(GS_{P1(S)}) = IG(GS_{P2(S)}) = \frac{7}{16}$. This can bring the operator to think that the two projections are equal, but this is not the case. This can be noticed by observing the two granular structures $GS_{P1(S)}$ and $GS_{P2(S)}$ in Figure 7.4 that contains different elements in the coarser granule at the middle level. But this difference can be also measured by using the distance between granular structures (Eq. (7.6)). Indeed, although the distance between the current situation and the two projections is the same, i.e. $D(GS_S, GS_{P1(S)}) = D(GS_S, GS_{P2(S)}) = \frac{3}{16}$, the distance between the two projections is significant: $D(GS_{P1(S)}, GS_{P2(S)}) = \frac{6}{16} \neq 0$. Thus, when different granular structures have the same information granularity, the way to evaluate their dissimilarity is to consider the distance between them. Often the information granularity of the granular structures in a given SA application does not change over the time since the granules and the levels of granulation are strongly related with the hierarchies defined in the GDTA. In such cases, the distance between the structures appears the main indicator to measure the dissimilarity between situations.

The information granularity and the distance between two granular structures support comprehension and projection as they give measures of how much two situations are different. The human operator may consider such measures as early indicators of informative and structural differences between a recognized situation and a possible projection.

7.2.2 Dealing with Operator's Expectations: Conformity Analysis

Information granularity and distance are two measures that give structural information on the difference between situations. However, such measures give no insights about the interpretation of the situation or about a possible classification of the situation (e.g., bad, good, as expected, etc.). To provide the operators also with such kind of information, we define a conformity analysis approach to understand if a recognized situation conforms to the expectations of an SA operator. In Section 2.1, we evidenced that the expectations of the human operators play a major role in fastening the process of situation assessment and decision making, since when operators expect a particular situation, they do not need to rely on their mental model, on the working memory and on the other cognitive mechanisms to process further data and identify the situation. Obviously, wrong expectations cause a critical loss of SA with bad effects on the decision making process. The conformity analysis, via the verification of operators' expectations with the current situation (modeled with granular structure), is useful to confirm or disprove such expectations in a short time. The idea on which is based the conformity analysis is to provide a linguistic description of granules in a granular structure and then compare these descriptions with a set of expectations formalized with fuzzy if-then rules.

Conformity analysis is inspired by the fuzzy pattern matching technique proposed by [274]. For this reason, we present such approach by considering the use of a centroid-based clustering technique for the granulation process to obtain granular structures, and then we describe the obtained clusters as fuzzy patterns (thus using linguistic variables), on which we apply the conformity analysis. However, the approach can be generalized to other formalisms that allow to create granular structures starting from raw data and to describe them with some kind of linguistic pattern.

The granules are defined per spatial proximity by using a centroid-based clustering technique. Consequently, a granule is a cluster of observations that are close in the space and can be described as a fuzzy pattern in the form x_1 *is* A_1 AND ... x_k *is* A_k where x_k is the k-th attribute of the centroid of the cluster and A_k is a family of linguistic variables. We can also evaluate a confidence degree of the linguistic description as a t-norm $\sigma = min(\mu_{A_1}, ..., \mu_{A_k})$. σ gives information on the 'strength' (i.e., the level of confidence, of reliability) of the linguistic description associated to a granule. Each pattern can be classified with respect to some criterion related with the specific problem (e.g., good/bad; safe/unsafe; low/medium/high, etc.). Such classification of the pattern can be considered as an association rule of the form

$$R_i : \ x_{i,1} \ is \ A_{i,1} \ AND \ ... \ x_{i,k} \ is \ A_{i,k} \ \rightarrow \ Class \ is \ C_i \tag{7.9}$$

where C_i can be a fuzzy set or a categorical value. If we can not classify the pattern with the available information on the granule, we should not assume anything on its classification.

A granular structure is the union of all the granules and thus can be succinctly formalized as:

$$R \ = \ \cup_i \ R_i \tag{7.10}$$

where $i \in [1, n]$ with n the number of granules of the structure.

We can represent the expectations of the human operator with the same approach based on the fuzzy patterns. In this case, the patterns are defined by exploiting the knowledge of the operator or of an expert, and are represented as follows:

$$E_j : \ y_{j,1} \ is \ B_{j,1} \ AND \ ... \ y_{j,k} \ is \ B_{j,k} \ \rightarrow \ Class \ is \ C_j \tag{7.11}$$

where y are attributes, B linguistic variables, and C fuzzy sets or categorical values. The set of expected rules is:

$$E \ = \ \cup_j \ E_j \tag{7.12}$$

and usually $j < i$, since experts do not provide a high number of rules.

We can rank a granular structure with respect to a set of expectations of an operator comparing 7.10 and 7.12.

Specifically, it is possible to compute the weights $w_{i,j}$ between R_i and E_j in order to rank the discovered patterns with regards to conformity and unexpectedness [274]. Unexpectedness refers to any kind of deviation with respect to expectations. In [274] the weights $w_{(i,j)}$ are computed in two phases:

- evaluating a degree of matching between the attributes names of R_i and E_j. This degree is evaluated using the following equation:

$$L_{i,j} \ = \ \frac{|A_{(i,j)}|}{\max(|e_j|, |r_i|)} \tag{7.13}$$

where $A_{(i,j)}$ is the set of attribute names that are common to the conditional parts of R_i and E_j, and $|e_j|$, $|r_i|$ are the numbers of attribute names in the conditional parts of respectively E_j and R_i. For the consequential parts, we suppose the name of the class is the same so it does not account in 7.13.

- Evaluating the degrees of matching between the attribute values with the degree of matching between the k-th attribute value of conditional parts indicated by $V_{(i,j)k}$ and the degree of value match of consequential parts indicated by $Z_{(i,j)}$. If we do not have the correct classification, we consider as 1 the degree of value matches of the consequential parts.

$V_{(i,j)k}$ and $Z_{(i,j)}$ can be calculated in different ways. In [274] different cases to evaluate similarity between attribute values are presented that depend also on the specific operators $\{< \; > \; = \; \neq \; ...\}$ involved in the rules. In our case, attribute values are represented with fuzzy sets in both R_i and E_j, so we need to find a similarity between two fuzzy sets and we can do this via the mutual subsethood [275]. Given two fuzzy sets A and B, the mutual subsethood measures the extent to which A equals B and can be evaluated via:

$$\varepsilon(A, B) = \frac{|A \cap B|}{|A| \; + \; |B| \; - \; |A \cap B|} \tag{7.14}$$

where $|.|$ is the cardinality of the fuzzy set. In the case of Gaussian membership function, $|A| = \int_{-inf}^{+inf} a(x) \;\; = \int_{-inf}^{+inf} \exp^{-(\frac{x-c}{\sigma})^2}$, and $|A \cap B|$ can be easily calculated based on the crossover points. Let us take a look at Fig. 7.5 from [6] reporting an example of mutual subsethood for two Gaussian membership functions, with $c_1 > c_2$ and $\sigma_1 > \sigma_2$. The crossover points are evaluated as follows:

$$h_1 = \frac{c_1 + \frac{\sigma_1}{\sigma_2} c_2}{1 + \frac{\sigma_1}{\sigma_2}} \tag{7.15}$$

$$h_2 = \frac{c_1 - \frac{\sigma_1}{\sigma_2} c_2}{1 - \frac{\sigma_1}{\sigma_2}} \tag{7.16}$$

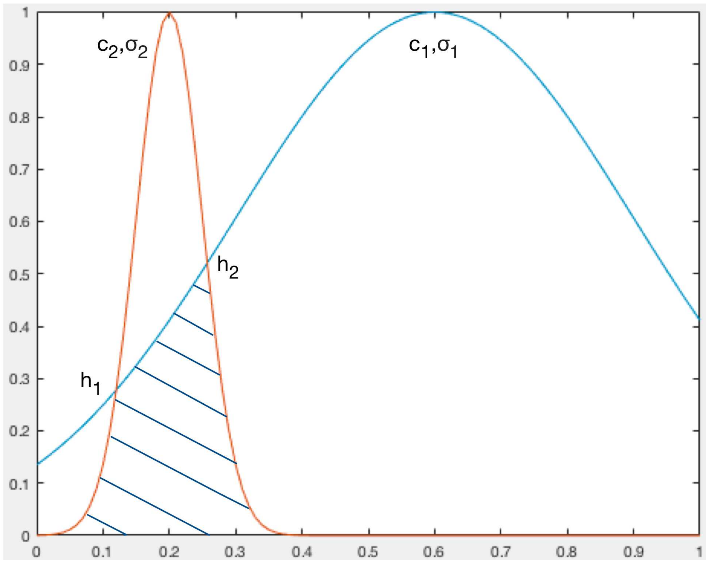


Figure 7.5 Mutual subsethood (our re-elaboration from [6])

Eq. 7.15 and 7.16 are used to calculate $|A \cap B|$ in all the cases for which $c_1 \neq c_2$. If $c_1 = c_2$ there are not crossover points and $h_1 = h2 = c$. In this case $|A \cap B| = \min(\sigma_1, \sigma_2)\sqrt{\pi}$. Details on the formulas, which are Gaussian integrals, to evaluate $|A \cap B|$ in the other cases are available in literature, for instance in annex of [6].

On the basis of $L_{(i,j)}$, $V_{(i,j)k}$ and $Z_{(i,j)}$, the weights $w_{(i,j)}$ for the **conformity** ranking can be calculated in the following way:

$$w_{(i,j)} = \begin{cases} \frac{Z_{(i,j)} * L_{(i,j)} * \sum_{k \in A_{(i,j)}} V_{(i,j)k}}{|A_{(i,j)}|} & \text{if } |A_{(i,j)}| \neq 0 \\ 0 & \text{if } |A_{(i,j)}| = 0 \end{cases} \tag{7.17}$$

The degree of match of a rule $R_i \in R$ with respect to the set of expected rules $E_j \in E$ is defined as:

$$W_i = \max(w_{(i,1)},\ w_{(i,2)},\ \ldots\ w_{(i,j)}) \tag{7.18}$$

If E is a set of expected rules by a human operator, then W_i represents a degree of conformity of a granule within a granular structure with respect to the expectations of the operator. This represents a degree of conformity of perceived information (represented as a granule) characterizing a recognized situation (represented as a granular structure) with the information expected (SA expectations) by the operators.

7.3 Evaluation

The objectives of our evaluation are related to a preliminary assessment of some of the benefits we envision for GrC in SA, specifically: to support comprehension and projection, and reduce L2 errors. The term *preliminary* here indicates the fact that we do not use a methodology for assessment of situation awareness, such as SAGAT [19], in a real scenario with real operators. We have implemented the proposed framework in a demonstrator using a clustering technique and used a synthetic data set to simulate an evolving situation, and evaluated how granular structures can be used to reason on evolving situations.

7.3.1 Using SOM to create Granules and Granular Structures

To create granules and granular structures we decided to use Self-Organizing Map (SOM) as a clustering technique. The Kohonen SOM [276] is an unsupervised neural network method particularly useful for data exploration and discovery of novel inputs. A SOM performs a topology-preserving mapping of the input data to the output units, enabling a reduction in the dimensionality of the input. This aspect gives SOM an added value related to visualization and visual inspections of the formed clusters. A SOM learns in a competitive way: the output neurons compete for the classification of the input patterns that are presented in the training phase. The output neuron with the nearest weight vector is classified as the winner. An output neuron is activated according to $Out_j = F_{min} \sum_i (x_i - w_{ji})^2$, where F_{min} is a threshold function, and w_{ji} is a connection weight between nodes j and i. Several works have compared SOM with other clustering techniques. In [277] SOM has been compared with k-means, fuzzy c-means and hierarchical clustering. Results show that SOMs generally have lower performance and are very sensitive to input data structure. In deciding to use SOMs for granulation in SA we accept a trade-off: to pay the cost of non-optimal clusters formation in favor of intuitive visualization features and easy data/pattern exploration that can offer benefits for SA.

7.3.2 Evaluation Scenario

To evaluate our approach, we refer to the already introduced surveillance scenario devoted to recognize anomalous situations, such as a splitting manoeuver. Let us suppose that two out of the four aircrafts are approaching the destination and the SA operator has to assess if they are proceeding close. Latitude and Longitude of the objects are mapped on a bi-dimensional area that has to be monitored by the operator. The normal situation is defined by a trajectory that the aircraft objects have to follow in approaching the destination. Fig. 7.6 shows an example of the normal trajectory from A to B that two aircrafts (depicted with red and blue points) have to follow in approaching B. The normal situation is when both the objects are approximatively in the area marked with two straight lines. When two objects are close and, at a certain time, one of the two suddenly changes, there is a split situation that is circled with an ellipse in the figure. In our scenario, to reason with granules, we have to induce a partition of the area under observation of Fig. 7.6 in several sub-areas that can group together per proximity the objects under surveillance. Fig. 7.7 shows a partition in 9 sub-areas of proximity that can be induced using three fuzzy sets and linguistic labels on the x and y dimensions of the area. The 9 partitions can be classified with respect to normal (N) or anomalous (A) positions that the objects can take in the area under observation. For each axis, we used Gaussian functions centred at 0, 0.5 and 1, with variance 0.175.

Starting from a data set of observations $o_j = (x_j, y_j)$ of posi-

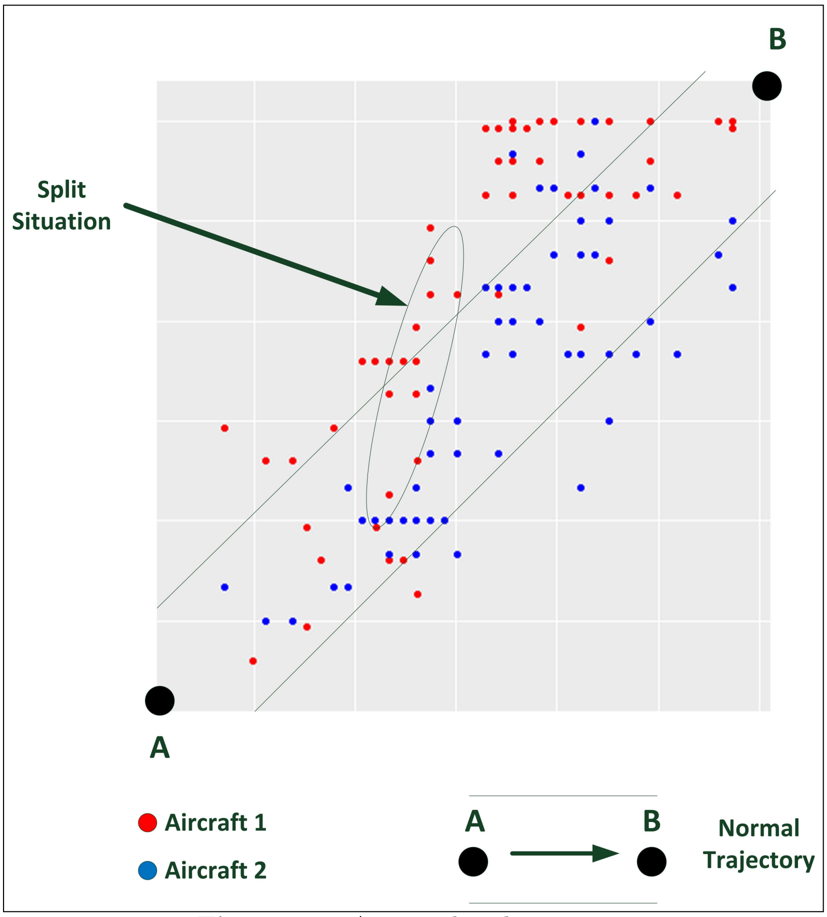


Figure 7.6 Area under observation

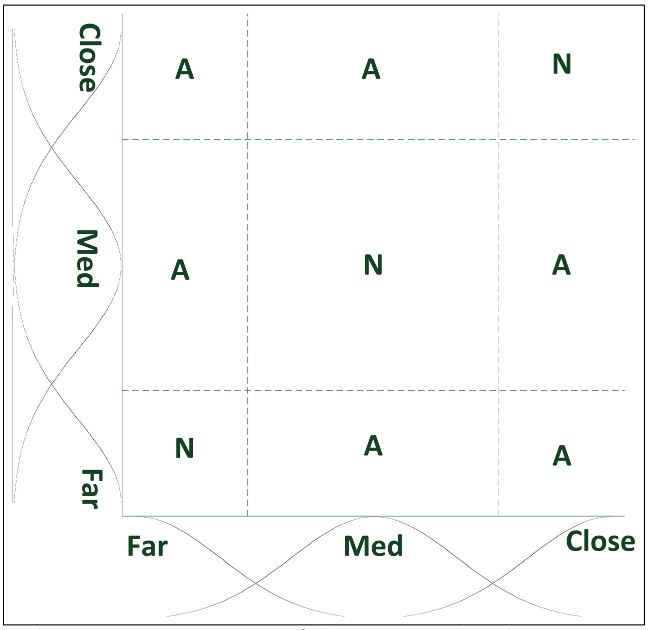


Figure 7.7 Partition of the area under observation

tions of an aircraft, we can create granules $g = \{o_j, o_k | o_j \approx o_k\}$ where $\approx$ is a proximity relation. A granule g groups a set of observations that are close together.

7.3.3 Definition of the Granular Structure

To create the granular structure we fuse granules g for aircrafts under observations. In our example, we limit to two objects and use a 3 x 2 SOM (where x refers to the multiplication sign) to fuse positions of the two objects. To train the SOM we use the data set that has been graphically shown in Fig. 7.6 and is representative of an evolutionary situation (the two objects are moving towards the destination) that includes a split maneuver. The trained map is shown in Fig. 7.8. Each neuron of the map is a granular structure fusing the position of two objects and the figure shows also the situations associated with the granular structures.

As we can see there are three granular structures that represent the situation GS2. The other three situations are represented

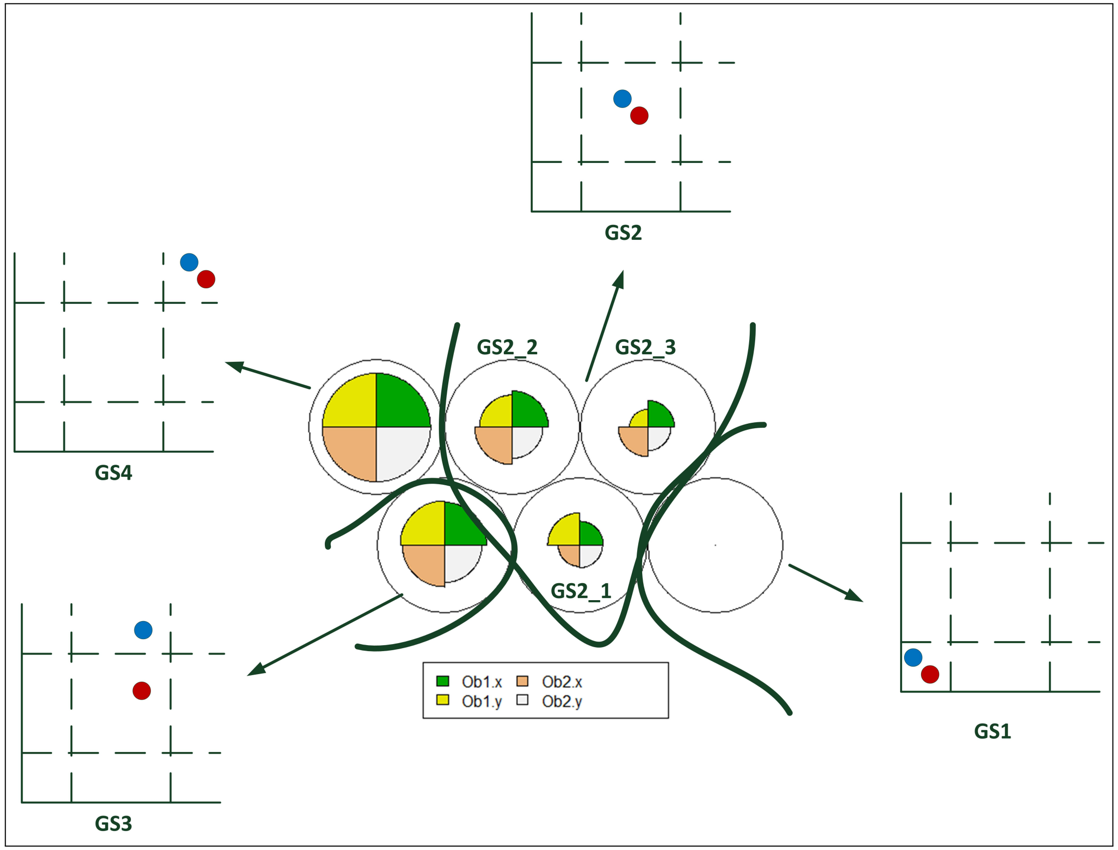


Figure 7.8 Granular Structures created with the SOM

with different granular structures. This is due to the fact that the partition we have done is larger in the middle of the area. Let us provide a linguistic description of the granular structures in the map and evaluate the conformity with respect to the expectations of the SA operators that can be succinctly described as $Ob_1\, is\, N\, AND\, Ob_2\, is\, N$, i.e. the two aircrafts move close together along a normal trajectory. The set of expectations can be formalized as follows, where FAR, MED and CLO are fuzzy sets with Gaussian membership functions previously reported:

$$Ob1.x\ is\ FAR\ and\ Ob1.y\ is\ FAR\ and\ Ob2.y\ is\ FAR\ and\ Ob2.y\ is\ FAR$$
$$Ob1.x\ is\ MED\ and\ Ob1.y\ is\ MED\ and\ Ob2.y\ is\ MED\ and\ Ob2.y\ is\ MED$$
$$Ob1.x\ is\ CLO\ and\ Ob1.y\ is\ CLO\ and\ Ob2.y\ is\ CLO\ and\ Ob2.y\ is\ CLO$$

The results are shown in Table 7.1 that reports the linguistic description of each granular structure, the confidence degree σ associated to the description, and the rank of conformance with the set of expectations evaluated with (10) and (11). We report as separated the three different granular structures associated with the situation $GS2$, i.e. $GS2_1$, $GS2_3$ and $GS2_3$.

Table 7.1 Linguistic interpretation of GS and ranking with expectations

GS	Ob1.x	Ob1.y	Ob2.x	Ob2.y	Confidence	Rank
$GS3$	Medium	Close	Medium	Medium	0.52789	0.75336
$GS2_1$	Medium	Medium	Medium	Medium	0.73895	1
$GS1$	Far	Far	Far	Far	0.89752	1
$GS4$	Close	Close	Close	Close	0.49204	1
$GS2_2$	Medium	Medium	Medium	Medium	0.83914	1
$GS2_3$	Medium	Medium	Medium	Medium	0.70742	1

As mentioned, a SOM map allows visual inspection of the data. In Fig. 7.8 we used a fan diagram style where, for all the granular structures created, the size of each variable for the two objects is clearly understandable. Another advantage of using SOM for spatial granulation is that the position of the neurons reflects proximity in the data. This means that granular structures positioned around a neuron represent probable projections of the granular structure represented by the neuron. The data set for our scenario, in fact, reports observations of a spatio-temporal evolution of the two objects. This has allowed us to replicate a case similar to the spatio-temporal granulation shown in Fig. 7.2, with the four granular structures resembling the case of an evolvable granular structure. This simplification is useful for our objectives that are to shows the benefits of reasoning with granular structures.

7.3.4 Reasoning with the Granular Structures

Now let us monitor the positions of ob_1 and ob_2 during three time windows. Table 7.2 reports the observations for the two objects in [t1, t2], and the associated granular structures of the map, i.e. $GS1$. If we review Table 7.1, an SA operator processing this information can easy perceive that the two objects are in a proximity region that is quite far from the destination. This information has an associated degree of confidence sufficiently high. At comprehension level, the operator recognizes the situation in this time window as in line with expectations, i.e. rank is 1. In summary, Table 7.2 reports a set of observations for which no anomalous situations are recognized. Also in terms of projections, looking at the neighbors of $GS1$, that are $GS2_1$ and $GS2_3$, an SA operator does not expect changes in the situations. The indicators of information granularity IG of the current situation $GS1$ and of its probable projections are the same, and the distance D between $GS1$ and its probable projections is zero. This means that the situations are similar from an informative and structural perspective. Also the ranking of the projections with respect to the expectations is 1 meaning the projected situations conform to the expectations.

Table 7.3 reports the observations for the two objects in another time window, i.e. [t4, t5], and the associated granular structures. Similar arguments can be provided for perception and comprehension levels. However, in this case a SA operator can receive early warnings on one of the probable projections of the situation recognized with the last observations. In fact, if we evaluate the IG of $GS2_2$ and of its projection $GS3$, we can see that are different and their distance is not zero. This indicates that the

Table 7.2 Observations in $t \in [t1, t2]$ and associated G

Ob1.x	Ob1.y	Ob2.x	Ob2.y	GS
0.02	0.00	0.02	0.00	GS1
0.00	0.04	0.05	0.04	GS1
0.02	0.04	0.07	0.04	GS1
0.02	0.04	0.07	0.04	GS1
0.02	0.00	0.07	0.00	GS1
0.03	0.08	0.08	0.12	GS1
0.03	0.04	0.08	0.04	GS1
0.03	0.04	0.08	0.04	GS1
0.03	0.04	0.08	0.04	GS1
0.05	0.04	0.10	0.04	GS1
0.07	0.12	0.12	0.17	GS1
0.10	0.04	0.15	0.04	GS1
0.10	0.08	0.15	0.12	GS1

situation is changing in this projection, and also the ranking value indicates that this projection does not conform too much to the expectations.

Table 7.3 Observations in $t \in [t4, t5]$ and associated GS

Ob1.x	Ob1.y	Ob2.x	Ob2.y	GS
0.34	0.62	0.34	0.42	$GS2_1$
0.44	0.37	0.49	0.42	$GS2_3$
0.51	0.66	0.51	0.46	$GS2_2$
0.53	0.70	0.53	0.50	$GS2_2$
0.53	0.70	0.53	0.50	$GS2_2$
0.54	0.70	0.54	0.50	$GS2_2$
0.54	0.66	0.54	0.46	$GS2_2$
0.56	0.70	0.56	0.50	$GS2_2$
0.58	0.66	0.58	0.46	$GS2_2$

Table 7.4 lastly reports the observations for the two objects in [t5, t6]. In this case, as anticipated in the previous slice, the situation changes and is not fully conform to the expectations. A deeper look at the finer granules for the two objects clearly shows that object ob_1 is moving away from the normal trajectory.

Table 7.4 Observations in $t \in [t5, t6]$ and associated GS

Ob1.x	Ob1.y	Ob2.x	Ob2.y	GS
0.58	0.74	0.58	0.54	GS3
0.59	0.78	0.59	0.58	GS3
0.59	0.78	0.59	0.58	GS3
0.61	0.74	0.61	0.54	GS3
0.64	0.74	0.64	0.54	GS3
0.64	0.91	0.64	0.71	GS3
0.66	0.91	0.66	0.71	GS3
0.68	0.78	0.68	0.58	GS3

7.3.5 Discussion

The preliminary evaluation reported in this section was not devoted to show how to create good granular structures with SOM but, instead, used a SOM as a rapid way to create granular structures resembling a case of evolvable granular structures for an evolving situation. This allowed us to show the benefits of our approach for comprehension and projection. However, the study of evolvable granular structure for evolutionary situations needs further conceptual development we left for future works.

As an example, in general, projecting into a near future requires the capability to perceive and comprehend evolutions of a granular structure. Given a universe U, the number of granular structures we can create is limited by the number of partitions of the universe. Furthermore, in real cases not all the partitions of the universe can be admissible projections of a situation. Knowing the rules that govern phenomena under observation can help in selecting the granular structures that can be considered as admissible projections. Moreover, the projection of a situation may depend also on the actions executed by actors of the situation. Having a clear view of the actions that are admissible in a specific situation, can give a strong support to SA operators for the issues at level 3 SA.

7.4 Model of Situations based on Rough Sets

The theoretical model based on neighborhood systems is useful to represent situations as granular structures by drawing together objects and elements according to different relations by specifying the right *distance* function $D(x, y)$ in Eq. 7.1. Specifically, when the SA requirements are about the position of objects, it can be useful to use a distance as in Section 7.2.2 with the clustering techniques. It is also possible to use a similarity function for grouping similar objects or fuzzy relationships to obtain fuzzy sets (useful to consider partial/vague relationships between the objects). In the most general cases, a binary functional relationship can be defined to consider any kind of relation between two objects.

In this section, we propose a computational model of situations which specializes the theoretical model of neighborhood system using the Rough Sets formalism. In this case, an equivalence relation is used to create granules of indistinguishable objects with respect to some attributes. The idea is to represent situations by means of i) an *information table* that contains the main elements of the environment (i.e., objects) and their attributes and ii) a granular structure consisting of a *lattice* that graphically represents the information table according to different, user-defined, information fusion criteria. It supports the comprehension of the current situation as it provides the human operators with detailed insights on the current state of the environment by means of the information table and with an interactive approach for reasoning on the situations thanks to the lattice that evolves over the time. Such a lattice supports the reasoning on groups of similar or indistinguishable objects (thus reducing the number of elements the human operators need to observe) and it supports also the reasoning on future evolutions and projections of situations. In such sense, it can accelerate the decision making processes. Moreover, the model is flexible as it is possible to change the information fusion criteria for generating a different lattice (that provides a different viewpoint on data) even at run-time. It is also a formal model, as it is based on rough set theory, thus enabling the possibility to compute measures of similarity among lattice structures. Such measures are useful for quantifying the differences among situations at different time intervals. It can be exploited in the definition of interactive systems for SA, as human operators can change at runtime the set of attributes on which they want to focus on or they can change the information fusion criteria.

In what follows, we briefly report some theoretical notions about rough sets on which the proposed model is based.

7.4.1 Rough Sets

Rough set theory has been introduced by Pawlak [278]. Its main application is the vague description of items, taking into account the uncertainty and vagueness of data. Rough sets have been widely used in a plethora of applications like Data Mining [279], pattern recognition [280] and many others. Rough sets are usually adopted to approximate formally a set with a pair of sets which give the lower and the upper approximation of the original set. At the core of this formalism, there are the concepts of information system and indiscernibility (or indistinguishability) relation.

More formally, let $I = (U, A)$ be an information system, where U is a non-empty set of finite objects and A is a non-empty finite set of attributes such that $a : U \to V_a$ for every $a \in A$, where V_a is the value set of a (i.e. the set containing all the values that a can take). An information system can be represented with an information table that assigns a value $a(x)$ from V_a to each attribute a and object $x \in U$ [278, 281]. Given any subset of attributes, $E \subseteq A$, we can define an equivalence relation as:

$$IND(E) = \{(x, y) \in U \times U | \forall a \in E, a(x) = a(y)\} \tag{7.19}$$

$IND(E)$ states that x and y are indiscernible (or indistinguishable) by attributes from E. An equivalence relation can be defined based on a set of attributes in an information table so that two objects are equivalent if and only if they have the same value on every attribute.

Given an equivalence relation E, we can define an equivalence class:

$$[x]_E = \{y | y \in U,\ x\ E\ y\} \tag{7.20}$$

Suppose $H \subseteq U$ is a set of objects we want to describe, or approximate, with the equivalence classes. With rough sets we can approximate H by constructing its lower and upper approximations:

$$\underline{apr(H)} = \{x | x \in U,\ [x]_E \subseteq H\} \tag{7.21}$$

$$\overline{apr(H)} = \{x | x \in U,\ [x]_E \cap H \neq 0\} \tag{7.22}$$

The lower and upper approximation can be interpreted also in terms of three regions, as proposed by Yao in [7, 282]:

$$POS(H) = \underline{appr(H)} \tag{7.23}$$

$$NEG(H) = U - \overline{appr(H)} \tag{7.24}$$

$$BND(H) = \overline{appr(H)} - \underline{appr(H)} \tag{7.25}$$

7.4.2 Computational Approach for Situation Modeling and Reasoning based on Rough Sets

The overall approach for representing situations with granular structures introduced in Section 7.1 can be further specialized with the application of rough sets for modeling situations. Figure 7.9 depicts the approach based on rough sets for supporting SA at all the three levels.

The starting point is a GDTA [15] providing requirements for the three levels of SA. From SA level 1 requirements, we can define an information table reporting on the rows the objects/elements $O_1, ..., O_n$ to be perceived at level 1, and on the columns the attributes $a_1, ..., a_m$. We can also add a column with a decisional attribute to classify the level 1 objects but this is not reported

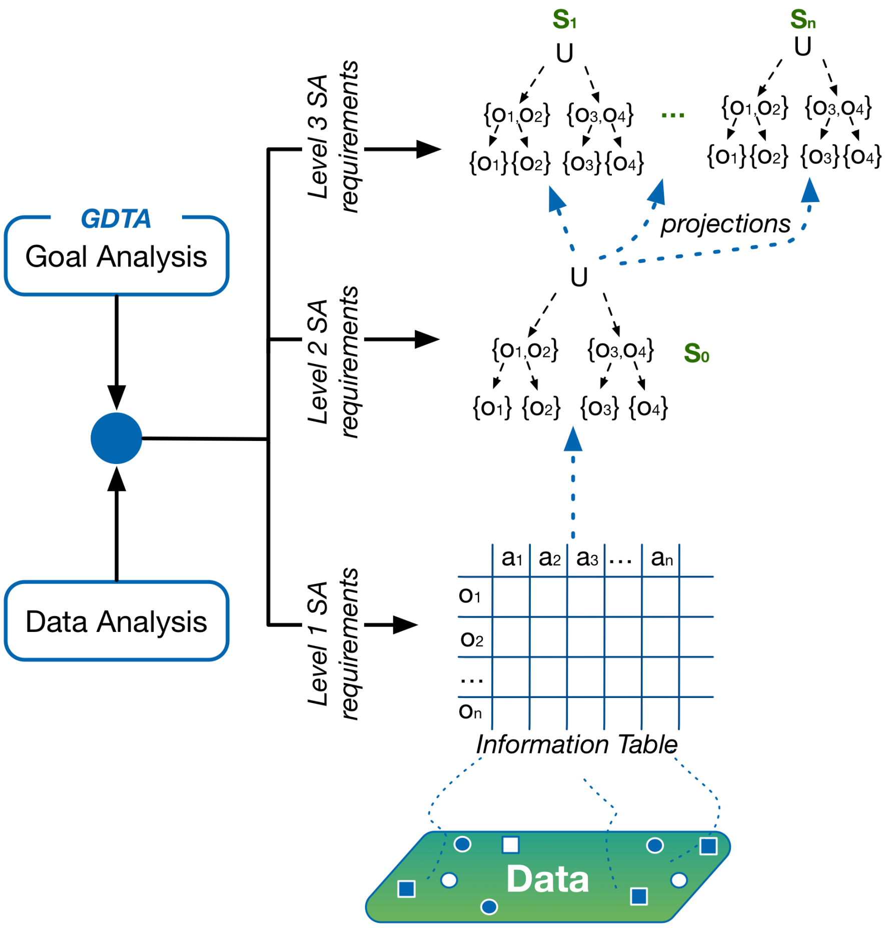


Figure 7.9 Computational approach based on Rough Sets for situation perception, comprehension and projection

in the figure. From this information table, we can group objects that are indistinguishable according to a subset of attributes by considering equation 7.1 with an equivalence relation instead of a distance function:

$$n(x) = \{y|(x,y) \in U \times U|\forall a \in E, a(x) = a(y)\} \tag{7.26}$$

which means that the neighborhood of x consists of all the objects for which it is valid the equivalence relation $IND(E)$ of Eq. 7.19. For instance, if E is the subset of attributes related to the velocity of an object, we can fuse objects that are equivalent with respect to speed. In this way, taking into account the constraints that SA level 2 requirements pose on the criteria to fuse information, we can define a lattice of partitions where each partition represents a group of objects that are equivalent with respect to a subset of attributes. For instance, S_0 represents a situation where objects $\{O_1, O_2\}$ and $\{O_3, O_4\}$ are equivalent with respect to B. This structure may be refined by considering different nested subsets of attributes. In fact, if we consider $C \subset B$ we may discern the four objects in S_0. It is worth mentioning that using incremental learning approaches, such as the one described in [283], the information table may be incrementally updated when attributes values vary over time, and this implies updating the corresponding lattices structures. If we consider all the subsets of attributes that can be also nested, e.g. $... \subset C \subset B \subset A$, it is possible to derive a set of lattices such as $S_1, ..., S_n$, and some of them can be possible evolutions of the recognized situation S_0. A human operator may have a good mental model and expertise to foresee possible evolutions starting from a recognized situation, such as S_0, but he/she may rather have some difficulties in reasoning on conceptual and informative differences among the possible evolutions and/or between a recognized situation and its evolution. The following subsections provide details on situation modeling with the rough sets formalism and on heuristics to support rapid reasoning on situation projections.

7.4.3 Situation model

A situation modeled with the rough sets is a combination of an *information table* IT and a lattice of partitions L over a subset of attributes A. Formally, we define the situation $S =< IT, L >$ where IT is an *information table* and L is a lattice of partitions defined over a sequence of nested attributes. Let be $F(x)$ a non-empty family of partitions (e.g. of equivalence classes), defined over a sequence of nested attributes, e.g. $A_3 \supset A_2 \supset A_1$. We define the equivalence relations on this sequence of subsets $I = E_{A_3} \subset E_{A_2} \subset E_{A_1} \subset E_0 = U \times U$. The union of these families for all the elements of an universe defines a lattice of partitions L:

$$L = \cup_{i=1}^{|U|} F(x_i) \tag{7.27}$$

The lattice groups objects with respect to a criteria of equivalence, usually embedded in SA level 2 requirements, thus obtaining objects grouped together in equivalence classes according to the (nested) subsets of attributes.

From the perspective of a human operator, the tuple $< IT, L >$ is more informative with respect to other situation models. In fact, besides having information on the attributes of all the objects of an environment, the human operator has a human-readable structure that gives information on groups of objects that are equivalent with respect to the criteria of his/her interest. As mentioned, if B is the subset of attributes related to trajectory or speed or other criteria, lattice L_B gives rapid information on the objects that are indistinguishable with respect to these attributes. The human operator can set the criteria of his/her preferences, giving rise to different subsets of attributes, and look at the equivalent objects.

Another interesting aspect of modeling situations with rough sets is the concept approximation, i.e. the possibility of approximating an unknown concept with a known concept, with the support of three regions. Let be $[x]_E$ a group of objects indistinguishable with respect to the subset E. Let us reason on the objects of an environment, and suppose we include in IT a decisional attribute a_d that allows to classify these objects with respect to a class (e.g. "safe/dangerous"). So, in this case, the situation model is $S =< IT \cup A_d, L_E >$, where A_d is the set of decisional attributes. Suppose $H \subseteq U$ is a subset of objects we want to describe, or approximate, with the equivalence classes. With rough sets we can approximate H by constructing its lower and upper approximations as described by Eq. (7.21) and Eq. (7.22) that can be also interpreted in terms of regions:

$$POS(H) = \underline{apr(H)} \tag{7.28}$$

$$NEG(H) = U \; - \; \overline{apr(H)} = \{x|x \in U, \; [x]_E \cap H = 0\} \tag{7.29}$$

$$BND(H) = \overline{apr(H)} \; - \; \underline{apr(H)} = \{x|x \in U, \; [x]_E \cap H \neq 0 \,, [x]_E \nsubseteq H\} \tag{7.30}$$

Eq. (7.28) is the positive region and includes all the equivalence classes that can be positively classified as belonging H, Eq. (7.29) is the negative region and includes objects that can be definitely ruled out as members of H and Eq. (7.30) is the boundary region consisting of objects that can neither be ruled in nor ruled out as members of the target set H.

This can be done also introducing a degree of tolerance as reported in [7]. We can introduce three-way decision rules, namely, positive rules for accepting an object as a member of H, negative rules for rejecting it, and boundary rules for deferring a definite decision. Let $P(H|[x]_E)$ be the conditional probability of an object belonging to H given that the object belongs to $[x]_E$. This probability can be estimated as

$$P(H|[x]_E) = \frac{|H \cap [x]_E|}{|[x]_E|} \tag{7.31}$$

where $|.|$ is the cardinality operator. If we consider probabilistic rough sets [7], a pair of thresholds α and β with $\alpha > \beta$ can be introduced, and by using the conditional probability defined in Eq. (7.31), the three regions in Eq. (7.28), Eq. (7.29) and Eq. (7.30) can be formulated as follows:

$$POS(H) = \{x|x \in U, \; P(H|[x]_E) \geq \alpha\} \tag{7.32}$$

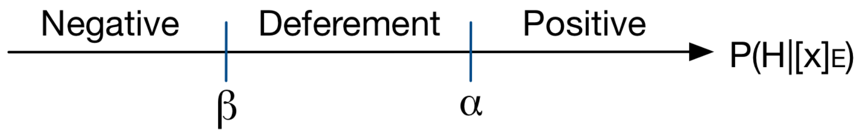


Figure 7.10 Probabilistic Rough Sets: regions of decisions ([7])

$$NEG(H) = \{x | x \in U,\ P(H|[x]_E) \leq \beta\} \tag{7.33}$$

$$BND(H) = \{x | x \in U,\ \beta < P(H|[x]_E) < \alpha\} \tag{7.34}$$

Figure 7.10 shows the regions we can define on the basis of probabilistic rough set model.

The value of these three regions for a human operator is easy to explain. Suppose H is the subset of objects that for a specific situation are classified as "safe" or "good". When a class of equivalence, i.e., a group of objects indistinguishable with respect to some criteria, is recognized in the current situation, with the support of Eq. (7.32), a human operator may know if this class can be approximated with the set of "good" or "safe" objects. This can improve the comprehension of the human operator.

The model of situation is useful to support the situation projections. An experienced human operator, leveraging on well developed mental models, may be able to project a given situation in the future in order to understand how this situation may evolve and what can be the differences. In order to help human operators (also those inexperienced), a dissimilarity and a similarity measures can be defined by adapting Eq. 7.6 and 7.8. Given two situations represented with the lattices of partitions L_1 and L_2, the dissimilarity between the two situations is:

$$Dis(L_1, L_2) = \frac{1}{|U|} \sum_{i=1}^{|U|} \frac{|L_1(x_i) \bigtriangleup L_2(x_i)|}{|U|} \tag{7.35}$$

where $|L_1(x_i) \bigtriangleup L_2(x_i)|$ is the cardinality of the symmetric difference between the family of partitions: $|F_1(x_i) \cup F_2(x_i)| - |F_1(x_i) \cap F_2(x_i)|$. The symmetric difference removes the common elements between two partitions, and can be considered as a sort of dissimilarity between the two structures. We can define the similarity as:

$$Sim(L_1, L_2) = 1 - D(L_1, L_2) \tag{7.36}$$

The dissimilarity (or the similarity) measure can be used for early evaluation of the projections of a situation.

Starting from the situation at time t_0, defined by the tuple $S_0 =< IT, L_B >$ over a subset $B \subset A$ of attributes, a human operator with a good mental model, or an automatic process, can foresee the evolution on some parameters of one or more objects, and update the information table IT. This will create evolutions of the current situation $S_1, ..., S_n$ that are possible projections of S_0. These projections may be evaluated with respect to S_0 by using Eq. (7.35) and Eq. (7.36) to understand if they differ or not. When needed, the human operator can also decide to change the subset of attributes for the creation of lattice structures representing projections, and this can be useful if the criteria behind level 3 SA requirements differ from the ones of SA level 2. Also in this case, Eq. (7.35) and Eq. (7.36) can give rapid information on differences with respect to the *status quo*.

Lastly, the partitions of the projected situations may be classified according to the three-way decision rules, to approximate the new classes of the projected situations with the known ones. This can help improving the comprehension of the projected situations.

7.5 A Case Study on monitoring Vessel Traffic

7.5.1 Scenario

We propose a case study related to the monitoring of dangerous movements of vessels and ships in order to demonstrate how the proposed model of situations based on rough sets can be helpful in supporting human operators to anticipate abnormal conditions and to be early warned of possible dangerous situations. In the maritime domain, it is crucial to understand why certain vessels

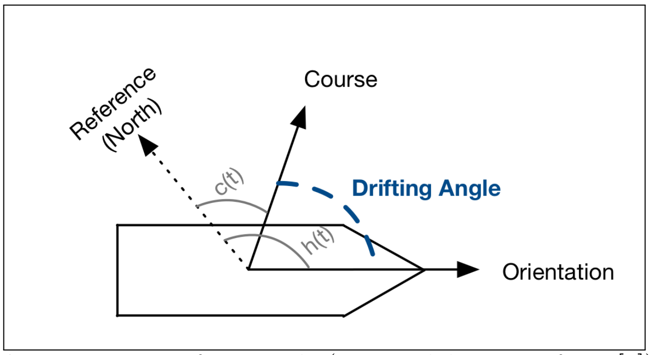


Figure 7.11 Drifting angle (our re-elaboration from [8])

movements take place. A surveillance operator, by observing the trajectories, the speed, the characteristics of the vessels, can identify dangerous situations and proceed to further investigations. In such situations, the human operator intervenes because he/she knows (i.e., he/she has the right mental model) the normal behavior of vessels in the observed environment, and so he/she can identify abnormal conditions [8]. But, in case of many vessels, different kind of ships and heavy traffic, a human operator may be not able to early identify dangerous and abnormal situations. Specifically, a hint about a dangerous situation is represented by the drifting movement of a vessel. A vessel may start to drift due to engine failure, that makes the vessel uncontrollable. A vessel is said to be drifting [8] when it is moving slowly, usually with a velocity $v(t)$ between 3 and 5 knots, and its course $c(t)$ and orientation $h(t)$ have a significant difference, usually more than 30°, as depicted in Figure 7.11.

We want to identify potential drifting vessels for supporting the human operator in understanding the movements of such vessels, so to early act for keeping safe the overall situation in the stretch of water under control.

Let us consider the information table in Table 7.5 that reports the values of a group of 6 vessels in an area under maritime surveillance by the human operator. The position and the drifting angles of the vessels are depicted in Fig. 7.12. In particular, we assume that the set of attributes A of the information table is $A =$

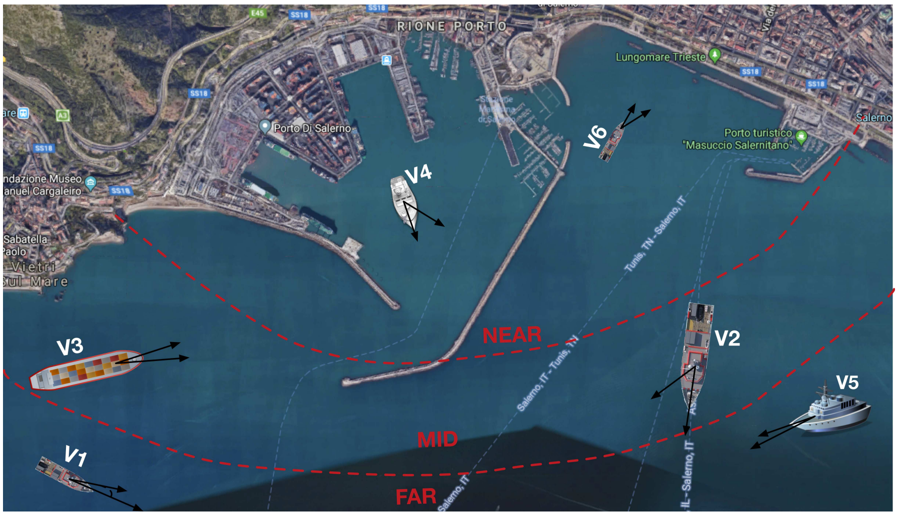


Figure 7.12 Position and drifting angles of the six vessels of the case study.

Table 7.5 Information table for the vessel traffic scenario

	Velocity	**Drifting Angle**	**Distance from coast**	**Type**	**Decision**
V1	LOW	LOW	FAR	Ferry	S
V2	MID	MID	MID	Cargo	S
V3	MID	LOW	MID	Cargo	S
V4	LOW	MID	NEAR	Ferry	D
V5	MID	LOW	FAR	Research	S
V6	LOW	LOW	NEAR	Ferry	S

$\{Velocity, DriftingAngle, DistancefromCoast, Type\}$, whose elements may assume the following values:

- Velocity: 1) LOW ($0\ knots < v(t) \leq 5\ knots$); 2) MID ($5\ knots < v(t) \leq 15 knots$); 3) HIGH ($v(t) > 15 knots$)
- Drifting Angle $[c(t) - h(t)]$: 1) LOW ($\leq 15°$); 2) MID ($> 15°$ and $\leq 30°$); 3) HIGH ($> 30°$)
- Distance from the coast: 1) NEAR ($\leq 2 miles$); 2) MID ($> 2 miles$ and $\leq 10 miles$; 3) FAR ($> 10 miles$)
- Type: 1) Cargo (commercial vessel); 2) Ferry (a ferry that usually moves between two points); 3) Research (vessel designed to perform research at sea).

The column *Decision* represents the decisional attribute that will be used in Section 7.5.3 for classifying the set of vessels: D stands for dangerous while S for safe. Such classification can be also obtained by exploiting the formalization of user's expectations, as discussed in Section 7.2.2.

7.5.2 Supporting Situation Comprehension

At time $t = t_0$, the subset of attributes the operator considers for performing the granulation consists of the attribute *Drifting Angle*, that is $B = \{DriftingAngle\} \subset A$, which leads to the following equivalence subclasses of vessels: {V1, V3, V5, V6} and {V2,V4}. The resulting lattice of partitions L_B is shown in Figure 7.13.A. By observing the granular structure, the human operator knows that the group of vessels consisting of {v2,v4} deserves attention because of the large drifting angle that may be an hint of a dangerous situation. To perform a deeper analysis, the human operator considers another attribute in the granulation process, and specifically he considers the *velocity* of the vessels. This means that now the sequence of nested attributes for partitioning the vessels is $C = \{Drifting\ Angle, Velocity\} \supset B = \{Drifting\ Angle\}$. In this way, the lattice of Fig. 7.13.A evolves in the lattice of Fig.

7.13.B, which still represents the same situation at time $t = t_0$ but considering further details. The new attribute allows to further partition the vessels: at the lower level of the lattice, a granule containing only the vessel $V4$ represents a potential drifting vessel since it has a large drifting angle with a low cruise velocity. The other vessel $V2$ has a higher velocity, and so, with the available information, it can not be considered a drifter because its engine works well. Thus, it is possible that the vessel $V2$ is doing a normal and safe maneuver. Obviously, the human operator can use other information, when available, in order to further improve the comprehension of what is happening in the environment. The human operator, by leveraging on his/her mental model and experience, on the basis of the information represented via the lattice and the information table, decides if the identified situation can be considered as still safe or it requires some actions to avoid accidents.

7.5.3 Improving Comprehension by classifying Objects with Probability Rough Sets

The decisional attribute is useful to automatically classify the state of a granule of vessels. The probability of each group of objects (in the lattice) belonging to the class of safe objects can be computed with Eq. 7.31. In this way, the human operator can exclude a set of objects from further investigations as they are classified as being safe, and he/she can concentrate his/her effort on a smaller set of vessels. The decisional attribute (shown in the last column of Table 7.5) can be obtained by employing some domain rules that reflects the expectations of the operators, as we have described in Section 7.2.2.

In Table 7.5, $S = Safe$ indicates that the current movement of the vessel can be considered as normal and safe, and $D = Dangerous$ indicates the opposite situation. Using such attribute and Eq. (7.32), Eq. (7.33) and Eq. (7.34), it is possible to enhance the lattice of Fig. 7.13.A with an indication of the dangerousness of each group of vessels. Let us consider the class

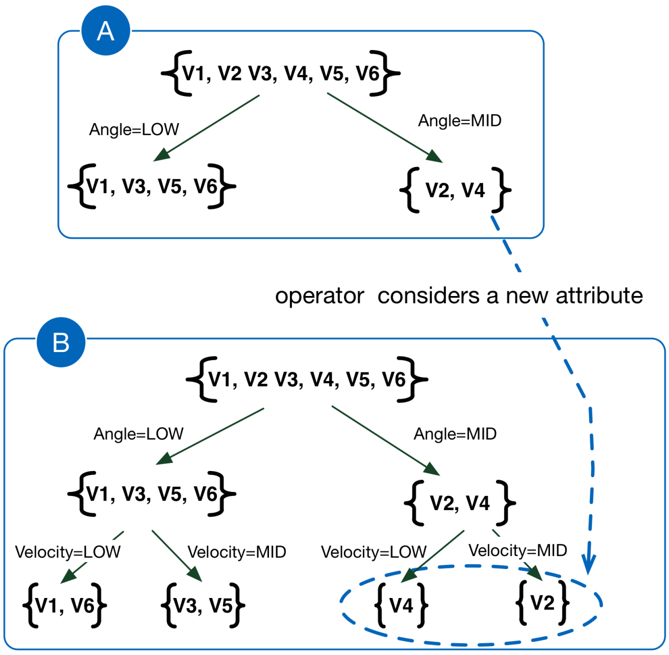


Figure 7.13 Supporting Comprehension with granular structures (lattices).

S of safe vessels. According to the information table 7.5, $S = \{V1, V2, V3, V5, V6\}$. At the lower level of the lattice depicted in Figure 7.13.A, related to the subset $B = \{DriftingAngle\} \subset A$ of attributes, there are two granules of vessels {V1,V3,V5,V6} and {V2,V4}. We want to classify these two subsets in the three regions of decision $POS(S)$ (containing groups of safe vessels), $NEG(S)$ (containing groups of not safe vessels), and $BND(S)$ (containing the group of vessels that can not be classified with the available information). By evaluating the conditional probability $P(S|[h]_B)$ of a vessel belonging to the class of $Safe$ objects S, we obtain: $P(S|\{V1, V3, V5, V6\}) = 1$ and $P(S|\{V2, V4\}) = 0.5$. Now, suppose that $\alpha = 0.63$ and $\beta = 0.25$ (such values are defined by Yao in [7]), we can classify the two subsets in this way: $POS(S) = \{V1, V3, V5, V6\}$, $BND(S) = \{V2, V4\}$ and $NEG(S) = \{\}$.

This means that the group of vessels $\{V1, V3, V5, V6\}$ is safe, while we defer the decision about the classification of the set $\{V2, V4\}$ when considering only the attribute *Drifting Angle*. By considering both the information on the drifting angle and on the velocity of the vessels (whose lattice is depicted in Fig. 7.13.B) we have: $P(S|\{V1, V6\}) = 1$ and $P(S|\{V3, V5\}) = 1$, $P(S|\{V2\}) = 1$, $P(S|\{V4\}) = 0$, and consequently: $POS(S) = \{\{V1, V6\}, \{V3, V5\}, \{V2\}\}$, $BND(S) = \{\}$, $NEG(S) = \{V4\}$. In this case, the human operator is aware that the vessel $V4$ needs particular attention as it is classified as a potential drifter, while he/she can leave out the other vessels from further investigations.

7.5.4 Supporting Situation Projection

What-if analysis can be used to reason on the possible evolutions of the current situation. In particular, starting from the lattice L_0 of Fig. 7.14.B which represents the situation of the vessels at time t_0 when considering both velocity and drifting angle, the human operator supposes that the angle of vessel $V4$ will increase in the near future. By applying the same sequences of nested attributes $\{Drifting\ Angle\} \subset \{Drifting\ Angle, Velocity\}$, and the new

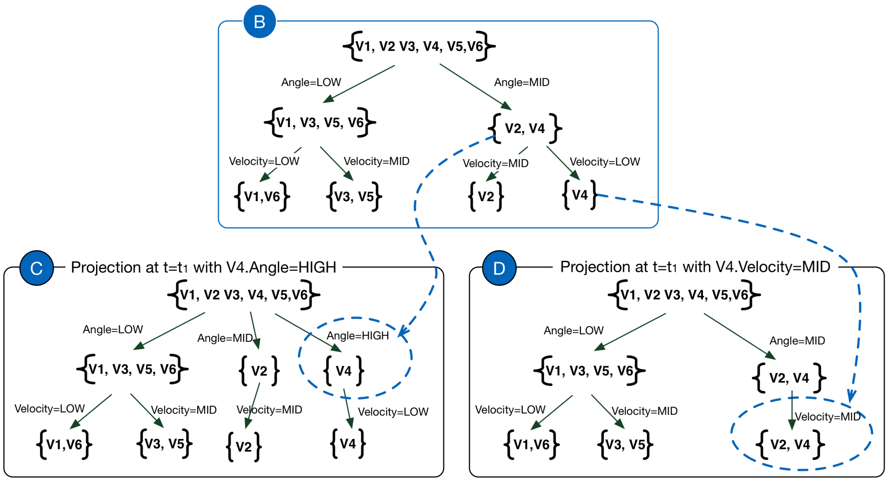


Figure 7.14 Supporting Projection with evolving lattices.

value for the *Drifting Angle* of $V4$, the lattice L_0 will evolve in the lattice L_1 of Fig. 7.14.C (the differences between the two lattices are circled in red).

By comparing the two lattices, the human operator observes that a new concept appears in the second lattice at the intermediate level of the hierarchy of granulation. Indeed, the subset $\{V2, V4\}$ is split in $\{V2\}$ and $\{V4\}$ due to the new value for the drifting angle. By using the dissimilarity function of Eq. (7.35), it is possible to quantify the differences between the two lattices and the related situations: $Dis(L_0, L_1) = 0.2$. This measure can be used in SA systems to send early warning to users about the modification of the current situation (i.e., change in the status quo). The human operator evaluates the new situation and he/she can easily understand that the vessel $V4$ can become a drifter as the drifting angle will be higher in the near future. Considering that the new projected situation differs from the previous one, he/she can perform some actions in order to maintain the same situation as the one at time $t = t_0$. The human operator can simulate also other scenarios. For instance, he/she can suppose that vessel $V4$ will not increase its angle, but it will increase its velocity. Another lattice will be obtained, as depicted in Fig. 7.14.D. The intermediate level of the hierarchy is not modified, while the two subsets $\{V2\}$ and $\{V4\}$ at the lower level are merged in one subset $\{V2, V4\}$. In this case, the situation becomes safer than the previous one, as vessel $V4$ increases its velocity and so it can not be a drifting vessel. This can be a situation more desirable than the previous one. Accordingly, the human operator can perform some actions (e.g., contacting the commander of vessel $V4$) in order to modify the current situation for obtaining the projected one. As described in section 7.5.3, if we have a decisional attribute associated with the projected situations, it is possible to classify each group of vessels as being safe or not. This allows us to evaluate the projected situations automatically, helping the human operator in deciding which can be the best action to perform (e.g., “Will the vessel Vi be in a safe position?”, “Can I maintain the current situation?”).

7.6 Summary

In this chapter, we presented a novel model for representing situations based on granular computing. The model consists of a set-theoretical framework based on neighborhood systems which allows to exploits different criteria for granulating objects that characterize the situation. Granules and granular structures represent the building blocks of this model of situations and enable reasoning mechanisms. Leveraging on this formal model, some measures of distance, similarity and dissimilarity have been defined to support all the three levels of SA. Moreover, the Conformity Analysis represents a way for reasoning on granular structures in order to reduce errors at comprehension level and biases, by adding information on how much a situation conforms to human expectations.

The overall set-theoretic framework has been specialized by using the rough set theory. The resulting model of situations gives the possibility to define an interactive approach for reasoning on situations, thus obtaining different perspectives of the environment, while supporting comprehension and projection.

The advantages of the proposed model rely on the formal representation of situations that allows to the machines to interpret situations and act according to them while providing also an interactive and human-readable way of understanding and projecting situations. Furthermore, it offers to human operators a high degree of flexibility that consists in the possibility of showing different perspectives of a situation by allowing the selection (in an interactive way) of different subsets of attributes to be used for creating the partitions of objects. The added values (for the human operators) of the situation model can be summarized as follows:

- improving perception and comprehension, via the provision of explicit information on the status of each element to be perceived and of a graphical, human-readable structure such as the lattice of partitions

- supporting the reasoning on different possibilities of forming partitions, by allowing him/her to identify different subsets of attributes that may match different SA level 2 criteria to fuse objects

- improving comprehension of the situations with concept approximation and classification, allowing a human operator to approximate the partitions of a lattice structure with known concepts

- checking human expectations via conformity analysis, in order to automatically verifying if the current situation conform with the expected one

- supporting rapid decision making with measures of dissimilarity between recognized and projected situations.

Chapter 8

Conclusion and Future Work

"Results! Why, man, I have gotten a lot of results.
I know several thousand things that won't work. "

— Thomas Edison

This chapter concludes the dissertation with a short summary and some final remarks on the main contributions of the research work. Moreover, some reflections on possible future work are discussed.

8.1 Summary

This study aims to define theoretical and practical approaches based on computational models and techniques in order to contribute to the resolution of the Situation Awareness (SA) demons. Specifically, the main idea of this research refers to the definition of computational approaches that are grounded on the cognitive mechanisms of humans to tackle with the common errors in SA caused by the SA demons.

A thorough investigation has been conducted on SA and GrC paradigms and on the state-of-the-art techniques, approaches and systems. Such an investigation allowed us to identify the most promising research areas that were worthwhile for further study to pursue our goals. In particular, we determined that data mining, ontologies and GrC are the three main areas to investigate for the definition of the computational approaches. The synergistic combination of data mining and ontologies have been identified as a valid approach to deal with issues at the perception level regarding the management of sensor data exacerbated by the demons of complexity and data overload. We defined an approach for the virtualization of physical sensors that enables a quality-aware management of sensor data, which includes a technique for data imputation based on association rule mining. Experiments on a real dataset showed that the proposed approach outperforms traditional data imputation techniques in the considered context.

Subsequently, we focused on ontologies and semantic approaches to define the Adaptive Goal-driven Situation Management (AGSM) approach to deal with attentional tunneling, data overload, wrong mental models, and misplaced salience. AGSM is able to support users in focusing on the right information. It prevents users to be overwhelmed by data while maintaining a global view of what is happening in the environment. The AGSM approach has been implemented in three prototypical systems in the following application domains: e-learning, fleet logistics and port logistics. In these application scenarios, experimental results highlight that AGSM can really improve the level of SA gained by the users; these results were obtained by using both the SAGAT methodology and a numerical simulation.

Lastly, we focused on the definition of a novel support to the SA by exploiting a paradigm that has not yet been thoroughly adopted in this field – that is, GrC. GrC has been used for the definition of a theoretical framework for approximate reasoning on situations and a model for representing situations, and it is capable of supporting the whole process of SA. These characteristics allowed us to tackle complex SA demons – such as wrong mental models, attentional tunneling, and complexity – in an innovative and effective way.

To the best of our knowledge, this is the first attempt at a thorough adoption of GrC – particularly of its principles, methodologies, and computational aspects – for improving SA systems. Although we are just at the beginning of the exploration of GrC techniques in the field of SA, our first proposals and experimentations already revealed the potentiality of this paradigm. In short, GrC represents a promising approach for data processing and decision making in SA, and it allows for an envisioning of a thorough, synergic and complete integration of the two paradigms.

8.1.1 Contributions

Pursuing the overall goal of addressing SA demons, besides the initial contributions related to a novel overview of SA research works and an original mapping between GrC and SA, in this thesis work we have proposed the following main contributions::

- the quality-aware sensor data management approach with the sensor data imputation technique based on association rule mining

- the Adaptive Goal-driven Situation Management approach, with the mechanism of goal selection based on goal desirability measure and the reinforcement learning technique for adapting this mechanism to the user's feedback

- the set-theoretical framework of GrC for approximate reasoning in SA

- the model of situations based on rough sets.

8.2 Final Remarks: Contributions to the Resolution of SA Demons

The contributions of this thesis were devoted to the resolution of the eight SA demons, as indicated in the following.

1. *Attentional tunneling*: (i) The AGSM approach supports users in the alternation between goal-driven and data-driven information processing, and it suggests which is the most important goal to attend to. This allows human operators to avoid remaining stuck in a goal and consequently losing the global SA. (ii) The framework of GrC and the model of situations support the human operator in avoiding attentional tunneling, due to the identification of early warnings that can be helpful to identify the most critical elements of the environment, which occurs independently of the current user goal.

2. *Data overload*: (i) The quality-aware sensor data management allows users to virtualize the sensors and to manage sensor data depending on the quality requirements. The virtual sensors can be configured to filter only relevant data and to provide only data that conforms to the desired quality. A virtual sensor can perform some low-level data fusion among the data gathered by multiple physical sensors. This helps in reducing the amount of data to observe; it also improves the perceived reliability by the users. (ii) The AGSM approach reduces the data overload problem by means of the automatic evaluation of events and alarm conditions to evaluate the desirability of goals, thereby relieving the workload on the users. (iii) the framework of GrC for SA reduced the data overload issue since it provides the users with a multilevel view of the environment. In this way, users can analyze a problem at the right level of granularity, without considering too many details when there is no need. Furthermore, the model of situations based on rough sets allows users to aggregate elements that are indistinguishable according to desired criteria, thereby reducing the number of single elements to consider

3. *Complexity*: The GrC framework deals with the complexity of the problem with the help of the multilevel structure representing the situations, which allows users to analyze data at the desired level of granularity. Moreover, the multiple perspectives from which a granular structure is analyzed can help in dealing with the complexity of the problem.

4. *Memory trap*: The model of situations based on rough sets

helps to relieve the workload on the working memory. Having a visual picture of the situations in the form of a granular structure allows users to directly see all the needed information in a single picture without the need to remember the states of different elements. Moreover, the multilevel view allows users to use a level of detail of information that is commensurate with the current situation. Coarser granules of information allow the workload on the memory to be reduced.

5. *Workload and stressors*: (i) The AGSM approach can reduce the workload since the automatic evaluation of goal desirability allows users to focus only on the information related on the current goal. (ii) The granular structure of the GrC framework can reduce the workload since users can look the data at a coarser level of granularity, and having such a level reduces the number of elements to attend to.

6. *Wrong mental models*: (i) The AGSM approach allows human operators to select the best mental model because it suggests the goal that needs the attention of the user. The correct configuration of the mental model depends also on the goal that is set by users. Moreover, the suggestions of the information that users should attend to helps the formation of a good mental model to deal with the current situation. (ii) Wrong mental models leads also to incorrect expectations. The GrC framework and the conformity analysis support the users in dealing with false expectations, and this often leads to the use of correct mental model for the interaction with the system. Moreover, the granular structure supports the understanding of the problem and thus the reinforcement of the correct mental model in specific situations.

7. *Misplaced salience*: (i) The AGSM approach – by means of the concept of goal desirability – helps the designers of SA systems to develop adaptive interfaces that put the salience on those elements of the environment only when this is really needed. (ii) The early warnings generated by means of the GrC framework for SA and the model of situations help to recall the attention of the users only on the critical information at a given moment. This can be used to design interfaces that avoid the problem of

misplaced salience.

8. *Out-of-the-loop*: (i) The AGSM approach supports users in being in the loop of monitoring and control of the system – as a result of the support to the alternation between goal-driven and data-driven information processing. Moreover, having notifications sent to users automatically when a goal becomes more desirable allows users to avoid being out of the loop due to an excessive automation of the system, as users need to rapidly recall their attention on what is happening. (ii) The early warnings generated by means of the situation model based on rough sets and the conformity analysis can be seen as tools to recall user attention on the system in the cases where too many automatisms put users out of the loop of control.

8.3 Future Work

In this section, we outline some directions for future research.

With regard to the quality-aware sensor data management approach, we plan to define an optimization technique to automatically identify the best precision factor to use when dealing with real values. We also considered other approaches for the discretization and representation of real values (e.g., by transforming and approximating them with fuzzy sets and rough sets). Moreover, incremental and online rule mining techniques should be explored for the definition of approaches, which may be used for updating the rule base when new data become available.

Regarding the AGSM approach, future work will deal with the automatic definition of desirability measures. We plan to define measures of desirability that are not based on expert knowledge but can be automatically learned or defined according to the possible future evolutions of the current situations (e.g., by using measures similar to the distance defined for the granular structures). In this way, a desirability measure can take into account the fact that users want to remain in the current situation or that they wish to reach a different target situation instead. In this case, the advantage is that the desirability measure is completely independent of the application domain. Moreover, the proposed approach should be extended to the Team and Collective Situation Awareness.

Lastly, regarding the set-theoretical framework for SA, future work aims at defining interactive dashboards for supporting the interaction of users with the granular structures representing situations. Moreover, we plan to explore other formalisms of GrC to model situations for creating a set of different situation models, with each one specialized for representing data with specific criteria.

Bibliography

[1] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human factors*, vol. 37, no. 1, pp. 32–64, 1995.

[2] Y. Yao, "A triarchic theory of granular computing," *Granular Computing*, vol. 1, no. 2, pp. 145–157, Jun 2016.

[3] S. Salehi, A. Selamat, and H. Fujita, "Systematic mapping study on granular computing," *Knowledge-Based Systems*, vol. 80, pp. 78–97, 2015.

[4] Intel. (2004) Intel lab data. [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html

[5] H. Estrada, O. Pastor, A. Martínez, and J. Torres-Jimenez, "Using a goal-refinement tree to obtain and refine organizational requirements," in *Computational Science and Its Applications – ICCSA 2004: International Conference, Assisi, Italy, May 14-17, 2004, Proceedings, Part IV.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 506–513.

[6] S. Paul and S. Kumar, "Subsethood based adaptive linguistic networks for pattern classification," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 33, no. 2, pp. 248–258, 2003.

[7] Y. Yao, "The superiority of three-way decisions in probabilistic rough set models," *Information Sciences*, vol. 181, no. 6, pp. 1080–1096, 2011.

[8] N. Willems, R. Scheepens, H. van de Wetering, and J. J. van Wijk, "Visualization of vessel traffic," in *Situation Awareness*

with Systems of Systems. New York, NY: Springer New York, 2013, pp. 73–87.

[9] M. R. Endsley and M. M. Robertson, "Training for situation awareness in individuals and teams," *Situation awareness analysis and measurement*, pp. 349–366, 2000.

[10] Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, *Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris.* BEA, 2012. [Online]. Available: https://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf

[11] International Nuclear Safety Advisory Group, *Summary Report on the Post-accident Review Meeting on the Chernobyl Accident*, ser. INSAG. Vienna: INTERNATIONAL ATOMIC ENERGY AGENCY, 1986, no. 1. [Online]. Available: http://www-pub.iaea.org/books/IAEABooks/3598/Summary-Report-on-the-Post-accident-Review-Meeting-on-the-Chernobyl-Accident

[12] The Joint Accident Investigation Commission of MV ESTONIA, *Final report on the MV ESTONIA disaster of 28 September 1994.* Helsinki: Edita Ltd, 1997. [Online]. Available: http://onse.fi/estonia/brindex.html

[13] M. Naderpour, S. Nazir, and J. Lu, "The role of situation awareness in accidents of large-scale technological systems," *Process Safety and Environmental Protection*, vol. 97, no. Supplement C, pp. 13 – 24, 2015, bhopal 30th Anniversary.

[14] M. R. Endsley, "A taxonomy of situation awareness errors," *Human factors in aviation operations*, vol. 3, no. 2, pp. 287–292, 1995.

[15] ——, *Designing for situation awareness: An approach to user-centered design.* CRC press, 2016.

[16] C. M. Schulz, V. Krautheim, A. Hackemann, M. Kreuzer, E. F. Kochs, and K. J. Wagner, "Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system," *BMC anesthesiology*, vol. 16, no. 1, p. 4, 2015.

[17] M. R. Grech, T. Horberry, and A. Smith, "Human error in maritime operations: Analyses of accident reports using the leximancer tool," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 46, no. 19. Sage Publications Sage CA: Los Angeles, CA, 2002, pp. 1718–1721.

[18] T. C. Stratmann and S. Boll, "Demon hunt - the role of endsley's demons of situation awareness in maritime accidents," in *Human-Centered and Error-Resilient Systems Development: IFIP WG 13.2/13.5 Joint Working Conference*. Cham: Springer International Publishing, 2016, pp. 203–212.

[19] M. R. Endsley, "Situation awareness global assessment technique (SAGAT)," in *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, May 1988, pp. 789–795 vol.3.

[20] "Summary of the various definitions of situation awareness," Royal Aeronautic Society, JAA ESSAI project, Tech. Rep. [Online]. Available: https://www.raes-hfg.com/crm/crm-reports.htm#proceeds

[21] L. Carroll and Air Force Research Lab Mesa AZ Warfighter Readiness Research Division, *Desperately seeking SA*. Defense Technical Information Center, 1992.

[22] T. J. Emerson, J. M. Reising, and H. G. Britten-Austin, "Workload and situation awareness in future aircraft," in *SAE Technical Paper*, no. No. 871803. SAE International, 10 1987.

[23] F. F. Haines and C. L. Flatau, *Night Flying*, 2nd ed., ser. Practical Flying Series. Tab Books, 1992.

[24] M. A. Dalrymple and S. G. Schiflett., "Measuring situational awareness of AWACS weapons directors," in *Situational Awareness in the Tactical Air Environment: Augmented Proceedings of the Naval Air Warfare Center's First Annual Symposium*, 1997.

[25] M. R. Endsley *et al.*, "Theoretical underpinnings of situation awareness: A critical review," *Situation awareness analysis and measurement*, pp. 3–32, 2000.

[26] M. R. Endsley, "Situation awareness misconceptions and misunderstandings," *Journal of Cognitive Engineering and Decision Making*, vol. 9, no. 1, pp. 4–32, 2015.

[27] A. Salfinger, W. Retschitzegger, and W. Schwinger, "Maintaining situation awareness over time – a survey on the evolution support of situation awareness systems," in *2013 Conference on Technologies and Applications of Artificial Intelligence*, Dec 2013, pp. 274–281.

[28] Y. Zhang, S. Huang, S. Guo, and J. Zhu, "Multi-sensor data fusion for cyber security situation awareness," *Procedia Environmental Sciences*, vol. 10, no. Part B, pp. 1029 – 1034, 2011, 2011 3rd International Conference on Environmental Science and Information Application Technology ESIAT 2011.

[29] G. Shafer *et al.*, *A mathematical theory of evidence.* Princeton university press Princeton, 1976, vol. 1.

[30] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang, "Sensor fusion using dempster-shafer theory [for context-aware hci]," in *IMTC/2002. Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference (IEEE Cat. No.00CH37276)*, vol. 1, 2002, pp. 7–12 vol.1.

[31] F. Hillen, B. Hofle, M. Ehlers, and P. Reinartz, "Information fusion infrastructure for remote-sensing and in-situ sensor data to model people dynamics," *International Journal of Image and Data Fusion*, vol. 5, no. 1, pp. 54–69, 2014.

[32] E. W. N. K. Sanjay K. Boddhu, Robert L. Williams, "Increasing situational awareness using smartphones," *Proc.SPIE*, vol. 8389, pp. 8389 – 8389 – 13, 2012.

[33] C. Thompson, J. White, B. Dougherty, A. Albright, and D. C. Schmidt, "Using smartphones to detect car accidents and provide situational awareness to emergency responders," in *Mobile Wireless Middleware, Operating Systems, and Applications: Third International Conference, Mobilware 2010, Chicago, IL, USA, June 30 - July 2, 2010.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 29–42.

[34] X. Su, H. Tong, and P. Ji, “Activity recognition with smartphone sensors,” *Tsinghua Science and Technology*, vol. 19, no. 3, pp. 235–249, June 2014.

[35] M. Shoaib, H. Scholten, and P. J. M. Havinga, “Towards physical activity recognition using smartphone sensors,” in *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*, Dec 2013, pp. 80–87.

[36] M. Naderpour and J. Lu, “Supporting situation awareness using neural network and expert system,” in *The 10th International FLINS Conference on Uncertainty Modeling in Knowledge Engineering and Decision Making.* World Scientific, 2012.

[37] A. Coronato, G. D. Pietro, and G. Paragliola, “A situation-aware system for the detection of motion disorders of patients with autism spectrum disorders,” *Expert Systems with Applications*, vol. 41, no. 17, pp. 7868 – 7877, 2014.

[38] L. Snidaro, I. Visentini, and K. Bryan, “Fusing uncertain knowledge and evidence for maritime situational awareness via markov logic networks,” *Information Fusion*, vol. 21, no. Supplement C, pp. 159 – 172, 2015.

[39] P. Barnaghi, F. Ganz, C. Henson, and A. Sheth, “Computing perception from sensor data,” in *2012 IEEE Sensors*, Oct 2012, pp. 1–4.

[40] A. Salfinger, W. Retschitzegger, W. Schwinger, and B. Pröll, “CrowdSA - towards adaptive and situation-driven crowd-sensing for disaster situation awareness,” in *2015 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision*, March 2015, pp. 14–20.

[41] A. Crooks, A. Croitoru, A. Stefanidis, and J. Radzikowski, “#earthquake: Twitter as a distributed sensor system,” *Transactions in GIS*, vol. 17, no. 1, pp. 124–147, 2013.

[42] M. Basu, S. Bandyopadhyay, and S. Ghosh, "Post disaster situation awareness and decision support through interactive crowdsourcing," *Procedia Engineering*, vol. 159, no. Supplement C, pp. 167 – 173, 2016, humanitarian Technology: Science, Systems and Global Impact 2016, HumTech2016.

[43] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1079–1088.

[44] H. M. Saleem, F. A. Zamal, and D. Ruths, "Tackling the challenges of situational awareness extraction in twitter with an adaptive approach," *Procedia Engineering*, vol. 107, no. Supplement C, pp. 301 – 311, 2015, humanitarian Technology: Science, Systems and Global Impact 2015, HumTech2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877705815010383

[45] S. M. George, W. Zhou, H. Chenji, M. Won, Y. O. Lee, A. Pazarloglou, R. Stoleru, and P. Barooah, "Distressnet: a wireless ad hoc and sensor network architecture for situation management in disaster response," *IEEE Communications Magazine*, vol. 48, no. 3, pp. 128–136, March 2010.

[46] J. Ye, S. Dobson, and S. McKeever, "Situation identification techniques in pervasive computing: A review," *Pervasive and Mobile Computing*, vol. 8, no. 1, pp. 36 – 66, 2012.

[47] J. McCarthy, "Situation calculus with concurrent events and narrative," 2000.

[48] H. Levesque, F. Pirri, and R. Reiter, "Foundations for the situation calculus," 1998.

[49] H. J. Levesque, R. Reiter, Y. LespÃ©rance, F. Lin, and R. B. Scherl, "GOLOG: A logic programming language for dynamic domains," *The Journal of Logic Programming*, vol. 31, no. 1, pp. 59 – 83, 1997, reasoning about Action and Change.

[50] B. Li and J. Lijima, "A survey on application of situation calculus in business information systems," in *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, Nov 2007, pp. 425–424.

[51] A. Coronato and G. De Pietro, "Situation awareness in applications of ambient assisted living for cognitive impaired people," *Mobile Networks and Applications*, vol. 18, no. 3, pp. 444–453, Jun 2013.

[52] J. Barwise and J. Perry, *Situations and Attitudes.* Cambridge, MA: MIT Press, 1983.

[53] K. Devlin, "Situation theory and situation semantics," *Handbook of the History of Logic*, vol. 7, pp. 601–664, 2006.

[54] S. Mechkour, "Overview of situation theory and its application in modeling context," in *Seminar paper*, 2007.

[55] R. Dapoigny and P. Barlatier, "Formal foundations for situation awareness based on dependent type theory," *Inf. Fusion*, vol. 14, no. 1, pp. 87–107, Jan. 2013.

[56] A. Ranganathan, J. Al-Muhtadi, and R. H. Campbell, "Reasoning about uncertain contexts in pervasive computing environments," *IEEE Pervasive Computing*, vol. 3, no. 2, pp. 62–70, April 2004.

[57] K. Henricksen and J. Indulska, "Developing context-aware pervasive computing applications: Models and approach," *Pervasive and Mobile Computing*, vol. 2, no. 1, pp. 37 – 64, 2006.

[58] S. W. Loke, "Representing and reasoning with situations for context-aware pervasive computing: A logic programming perspective," *Knowl. Eng. Rev.*, vol. 19, no. 3, pp. 213–233, Sep. 2004.

[59] ——, "Logic programming for context-aware pervasive computing: Language support, characterizing situations, and integration with the web," in *Web Intelligence, 2004. WI 2004. Proceedings. IEEE/WIC/ACM International Conference on*, Sept 2004, pp. 44–50.

[60] G. Shafer, "Dempster-shafer theory," *Encyclopedia of artificial intelligence*, pp. 330–331, 1992.

[61] S. McKeever, J. Ye, L. Coyle, and S. Dobson, "Using dempster-shafer theory of evidence for situation inference," in *Smart Sensing and Context: 4th European Conference, EuroSSC 2009, Guildford, UK, September 16-18, 2009. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 149–162.

[62] D. Zhang, J. Cao, J. Zhou, and M. Guo, "Extended dempster-shafer theory in context reasoning for ubiquitous computing environments," in *2009 International Conference on Computational Science and Engineering*, vol. 2, 2009, pp. 205–212.

[63] S. Mckeever, J. Ye, L. Coyle, C. Bleakley, and S. Dobson, "Activity recognition using temporal evidence theory," *J. Ambient Intell. Smart Environ.*, vol. 2, no. 3, pp. 253–269, Aug. 2010.

[64] A. Padovitz, S. W. Loke, and A. Zaslavsky, "Multiple-agent perspectives in reasoning about situations for context-aware pervasive computing systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 4, pp. 729–742, July 2008.

[65] A. Boytsov and A. Zaslavsky, "Formal verification of context and situation models in pervasive computing," *Pervasive Mob. Comput.*, vol. 9, no. 1, pp. 98–117, Feb. 2013.

[66] A. Boytsov, A. Zaslavsky, E. Eryilmaz, and S. Albayrak, "Situation awareness meets ontologies: A context spaces case study," in *Lecture Notes in Computer Science.* Springer, 2015.

[67] P. Delir Haghighi, S. Krishnaswamy, A. Zaslavsky, and M. M. Gaber, "Reasoning about context in uncertain pervasive computing environments," in *Smart Sensing and Context: Third European Conference, EuroSSC 2008, Zurich, Switzerland, October 29-31, 2008. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 112–125.

[68] C. Anagnostopoulos, Y. Ntarladimas, and S. Hadjiefthymiades, "Situational computing: An innovative architecture with imprecise reasoning," *Journal of Systems and Software*, vol. 80, no. 12, pp. 1993 – 2014, 2007, selected papers from the International Conference on Pervasive Services (ICPS 2006).

[69] J. Zhao, Y. Zhou, and L. Shuo, "A situation awareness model of system survivability based on variable fuzzy set," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 8, pp. 2239–2246, 2012.

[70] D. Furno, V. Loia, and M. Veniero, "A fuzzy cognitive situation awareness for airport security," *Control and Cybernetics*, vol. 39, no 4, pp. 960–982, 2010.

[71] R. E. T. Jones, E. S. Connors, and M. R. Endsley, "Incorporating the human analyst into the data fusion process by modeling situation awareness using fuzzy cognitive maps," in *2009 12th International Conference on Information Fusion*, July 2009, pp. 1265–1271.

[72] R. E. T. Jones, E. S. Connors, M. E. Mossey, J. R. Hyatt, N. J. Hansen, and M. R. Endsley, "Using fuzzy cognitive mapping techniques to model situation awareness for army infantry platoon leaders," *Computational and Mathematical Organization Theory*, vol. 17, no. 3, pp. 272–295, Sep 2011.

[73] S. Chandana, H. Leung, E. Bosse, and P. Valin, "Fuzzy cognitive map based situation assessment for coastal surveillance," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–6.

[74] F.-P. Pai, L.-J. Yang, and Y.-C. Chung, "Multi-layer ontology based information fusion for situation awareness," *Applied Intelligence*, vol. 46, no. 2, pp. 285–307, Mar 2017.

[75] M. M. Kokar, C. J. Matheus, and K. Baclawski, "Ontology-based situation awareness," *Information Fusion*, vol. 10, no. 1, pp. 83 – 98, 2009, special Issue on High-level Information Fusion and Situation Awareness.

[76] M. M. Kokar and M. R. Endsley, "Situation awareness and cognitive modeling," *IEEE Intelligent Systems*, vol. 27, no. 3, pp. 91–96, May 2012.

[77] Y. Dominguez, W. Nick, and A. Esterline, "Situations, identity, and the semantic web," in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, March 2016, pp. 109–115.

[78] A. Kayes, J. Han, and A. Colman, "An ontological framework for situation-aware access control of software services," *Information Systems*, vol. 53, no. Supplement C, pp. 253 – 277, 2015.

[79] R. Pearson, M. P. Donnelly, J. Liu, and L. Galway, "Generic application driven situation awareness via ontological situation recognition," in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, March 2016, pp. 131–137.

[80] G. Meditskos and I. Kompatsiaris, "iknow: Ontology-driven situational awareness for the recognition of activities of daily living," *Pervasive and Mobile Computing*, vol. 40, no. Supplement C, pp. 17 – 41, 2017.

[81] C. J. Matheus, M. M. Kokar, and K. Baclawski, "A core ontology for situation awareness," in *Proceedings of the Sixth International Conference on Information Fusion*, vol. 1, 2003, pp. 545–552.

[82] E. Miguelanez, P. Patron, K. E. Brown, Y. R. Petillot, and D. M. Lane, "Semantic knowledge-based framework to improve the situation awareness of autonomous underwater vehicles," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 5, pp. 759–773, May 2011.

[83] N. Baumgartner, W. Gottesheim, S. Mitsch, W. Retschitzegger, and W. Schwinger, "BeAware! situation awareness, the ontology-driven way," *Data & Knowledge Engineering*, vol. 69, no. 11, pp. 1181 – 1193, 2010, special issue on contribution of ontologies in designing advanced information systems.

[84] D. Riboni and C. Bettini, “Context-aware activity recognition through a combination of ontological and statistical reasoning,” in *Ubiquitous Intelligence and Computing: 6th International Conference, UIC 2009, Brisbane, Australia, July 7-9, 2009. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 39–53.

[85] T. Metzke, A. Rogge-Solti, A. Baumgrass, J. Mendling, and M. Weske, “Enabling semantic complex event processing in the domain of logistics,” in *Service-Oriented Computing – ICSOC 2013 Workshops: CCSA, CSB, PASCEB, SWESE, WESOA, and PhD Symposium, Berlin, Germany, December 2-5, 2013. Revised Selected Papers.* Cham: Springer International Publishing, 2014, pp. 419–431.

[86] N. Stojanovic and A. Artikis, “On complex event processing for real-time situational awareness,” in *Rule-Based Reasoning, Programming, and Applications: 5th International Symposium, RuleML 2011 – Europe, Barcelona, Spain, July 19-21, 2011. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 114–121.

[87] G. Vlahakis, D. Apostolou, and E. Kopanaki, “Enabling situation awareness with supply chain event management,” *Expert Systems with Applications*, vol. 93, no. Supplement C, pp. 86 – 103, 2018.

[88] T. Lu, X. Zha, and X. Zhao, “Multi-stage monitoring of abnormal situation based on complex event processing,” *Procedia Computer Science*, vol. 96, no. Supplement C, pp. 1361 – 1370, 2016, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 20th International Conference KES-2016.

[89] D. J. Patterson, L. Liao, D. Fox, and H. Kautz, “Inferring high-level behavior from low-level sensors,” in *UbiComp 2003: Ubiquitous Computing: 5th International Conference, Seattle, WA, USA, October 12-15, 2003. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 73–89.

[90] E. M. Tapia, S. S. Intille, and K. Larson, "Activity recognition in the home using simple and ubiquitous sensors," in *Pervasive Computing: Second International Conference, PERVASIVE 2004, Linz/Vienna, Austria, April 21-23, 2004. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 158–175.

[91] T. v. Kasteren and B. Krose, "Bayesian activity recognition in residence for elders," in *2007 3rd IET International Conference on Intelligent Environments*, Sept 2007, pp. 209–212.

[92] F. Johansson and G. Falkman, "Detection of vessel anomalies - a bayesian network approach," in *2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information*, Dec 2007, pp. 395–400.

[93] F. Mirmoeini and V. Krishnamurthy, "Reconfigurable bayesian networks for adaptive situation assessment in battlespace," in *Proceedings. 2005 IEEE Networking, Sensing and Control, 2005.*, March 2005, pp. 810–815.

[94] P. Wiggers, B. Mertens, and L. Rothkrantz, "Dynamic bayesian networks for situational awareness in the presence of noisy data," in *Proceedings of the 12th International Conference on Computer Systems and Technologies*, ser. CompSysTech '11. New York, NY, USA: ACM, 2011, pp. 411–416.

[95] C. Morales and S. Moral, "Regression methods applied to flight variables for situational awareness estimation using dynamic Bayesian networks," in *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, ser. Proceedings of Machine Learning Research, vol. 52. Lugano, Switzerland: PMLR, 06–09 Sep 2016, pp. 356–367.

[96] T. Damarla, "Hidden markov model as a framework for situational awareness," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–7.

[97] M. Andersson and G. Pettersson, "Improving situation awareness using aerial-mission recognition and temporal information," in

Proceedings of the 7th International Conference on Information Fusion, Stockholm, Sweden, 2004.

[98] P. Lison, C. Ehrler, and G.-J. M. Kruijff, "Belief modelling for situation awareness in human-robot interaction," in *RO-MAN*, 2010.

[99] S. Pournouri, B. Akhgar, and P. S. Bayerl, "Cyber attacks analysis using decision tree technique for improving cyber situational awareness," in *Global Security, Safety and Sustainability - The Security Challenges of the Connected World: 11th International Conference, ICGS3 2017, London, UK, January 18-20, 2017, Proceedings.* Cham: Springer International Publishing, 2016, pp. 155–172.

[100] S. Y. Lee and F. J. Lin, "Situation awareness in a smart home environment," in *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, Dec 2016, pp. 678–683.

[101] B. J. Rhodes, N. A. Bomberger, M. Seibert, and A. M. Waxman, "Maritime situation monitoring and awareness using learning mechanisms," in *MILCOM 2005 - 2005 IEEE Military Communications Conference*, Oct 2005, pp. 646–652 Vol. 1.

[102] N. Brannon, J. Seiffertt, T. Draelos, and D. Wunsch, "Coordinated machine learning and decision support for situation awareness," *Neural Networks*, vol. 22, no. 3, pp. 316 – 325, 2009, goal-Directed Neural Systems. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608009000367

[103] J.-Y. Yang, J.-S. Wang, and Y.-P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognition Letters*, vol. 29, no. 16, pp. 2213 – 2220, 2008.

[104] R. Li, M. Tapaswi, R. Liao, J. Jia, R. Urtasun, and S. Fidler, "Situation recognition with graph neural networks," *CoRR*, vol. abs/1708.04320, 2017.

[105] R. Ilin and L. Perlovsky, "Cognitively inspired neural network for recognition of situations," *International Journal of Natural Computing Research (IJNCR)*, vol. 1, no. 1, pp. 36–55, 2010.

[106] N. A. Bomberger, B. J. Rhodes, M. Seibert, and A. M. Waxman, "Associative learning of vessel motion patterns for maritime situation awareness," in *2006 9th International Conference on Information Fusion*, July 2006, pp. 1–8.

[107] H. Wang, X. Liu, J. Lai, and Y. Liang, "Network security situation awareness based on heterogeneous multi-sensor data fusion and neural network," in *Second International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)*, Aug 2007, pp. 352–359.

[108] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd, "At the flick of a switch: Detecting and classifying unique electrical events on the residential power line," in *UbiComp 2007: Ubiquitous Computing: 9th International Conference, UbiComp 2007, Innsbruck, Austria, September 16-19, 2007. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 271–288.

[109] T. Kanda, D. F. Glas, M. Shiomi, H. Ishiguro, and N. Hagita, "Who will be the customer?: A social robot that anticipates people's behavior from their trajectories," in *Proceedings of the 10th International Conference on Ubiquitous Computing*, ser. UbiComp '08. New York, NY, USA: ACM, 2008, pp. 380–389.

[110] X. W. Liu, H. Q. Wang, Y. Liang, and J. B. Lai, "Heterogeneous multi-sensor data fusion with multi-class support vector machines: Creating network security situation awareness," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 5, Aug 2007, pp. 2689–2694.

[111] X. Liu, H. Wang, J. Lai, and Y. Liang, "Network security situation awareness model based on heterogeneous multi-sensor data fusion," in *2007 22nd international symposium on computer and information sciences*, Nov 2007, pp. 1–6.

[112] J. Lu, B. Liu, G. Zhang, Z. Hao, and Y. Xiao, “A situation assessment approach using support vector machines as a learning tool,” *International Journal of Nuclear Knowledge Management*, vol. 3, no. 1, pp. 82–97, 2008.

[113] M. Riveiro, G. Falkman, and T. Ziemke, “Improving maritime anomaly detection and situation awareness through interactive visualization,” in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–8.

[114] J. Liu, X. W. Feng, J. Li, and D. X. Wang, “Cyber security situation awareness based on data mining,” in *Advanced Materials Research*, vol. 756. Trans Tech Publ, 2013, pp. 4336–4342.

[115] M. Perkowitz, M. Philipose, K. Fishkin, and D. J. Patterson, “Mining models of human activities from the web,” in *Proceedings of the 13th International Conference on World Wide Web*, ser. WWW ’04. New York, NY, USA: ACM, 2004, pp. 573–582.

[116] P. Palmes, H. K. Pung, T. Gu, W. Xue, and S. Chen, “Object relevance weight pattern mining for activity recognition and segmentation,” *Pervasive and Mobile Computing*, vol. 6, no. 1, pp. 43 – 57, 2010.

[117] S. Pournouri and B. Akhgar, “Improving cyber situational awareness through data mining and predictive analytic techniques,” in *Global Security, Safety and Sustainability: Tomorrow’s Challenges of Cyber Security: 10th International Conference, ICGS3 2015, London, UK, September 15-17, 2015. Proceedings.* Cham: Springer International Publishing, 2015, pp. 21–34.

[118] S. Mitsch, A. Muller, W. Retschitzegger, A. Salfinger, and W. Schwinger, “A survey on clustering techniques for situation awareness,” in *Web Technologies and Applications: 15th Asia-Pacific Web Conference, APWeb 2013, Sydney, Australia, April 4-6, 2013. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 815–826.

[119] C.-H. Chen, L. P. Khoo, Y. T. Chong, and X. F. Yin, “Knowledge discovery using genetic algorithm for maritime situational

awareness," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2742 – 2753, 2014.

[120] N. Dahal, O. Abuomar, R. King, and V. Madani, "Event stream processing for improved situational awareness in the smart grid," *Expert Systems with Applications*, vol. 42, no. 20, pp. 6853 – 6863, 2015.

[121] M. Zhang, B. H. Kang, and Q. Bai, "Association rule based situation awareness in web-based environmental monitoring systems," in *U- and E-Service, Science and Technology: International Conference UNESST 2010, Held as Part of the Future Generation Information Technology Conference, FGIT 2010, Jeju Island, Korea, December 13-15, 2010. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 224–232.

[122] K. K. Hewawasam, K. Premaratne, and M.-L. Shyu, "Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections," *Trans. Sys. Man Cyber. Part B*, vol. 37, no. 6, pp. 1446–1459, Dec. 2007.

[123] M. G. Cimino, B. Lazzerini, F. Marcelloni, and A. Ciaramella, "An adaptive rule-based approach for managing situation-awareness," *Expert Systems with Applications*, vol. 39, no. 12, pp. 10 796 – 10 811, 2012.

[124] M. Nilsson, J. van Laere, T. Ziemke, and J. Edlund, "Extracting rules from expert operators to support situation awareness in maritime surveillance," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–8.

[125] A. Carrio, C. Sampedro, A. Rodriguez-Ramos, and P. Campoy, "A review of deep learning methods and applications for unmanned aerial vehicles," *Journal of Sensors*, 2017.

[126] D. A. Noever and W. Regian, "Deep learning for cyber threat modeling and real-time situational awareness," in *National Fire Control Symposium, At Orlando Florida*, 2017.

[127] K. Tang and D. Crandall, "Applying deep learning to improve maritime situational awareness," in *Proceedings of 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2016)*, 2016.

[128] C. Y. Park, K. B. Laskey, P. C. G. Costa, and S. Matsumoto, "Multi-entity bayesian networks learning for hybrid variables in situation awareness," in *Proceedings of the 16th International Conference on Information Fusion*, July 2013, pp. 1894–1901.

[129] P. C. G. da Costa, K. B. Laskey, and K. Chang, "Prognos: applying probabilistic ontologies to distributed predictive situation assessment in naval operations," 2009.

[130] C. Y. Park, K. B. Laskey, P. C. G. Costa, and S. Matsumoto, "Predictive situation awareness reference model using multi-entity bayesian networks," in *17th International Conference on Information Fusion (FUSION)*, July 2014, pp. 1–8.

[131] N. Nwiabu, I. Allison, P. Holt, P. Lowit, and B. Oyeneyin, "Situation awareness in context-aware case-based decision support," in *2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, Feb 2011, pp. 9–16.

[132] B. Pavkovic, L. Berbakov, S. Vrane, and M. Milenkovic, "Situation awareness and decision support tools for response phase of emergency management: A short survey," in *2014 25th International Workshop on Database and Expert Systems Applications*, Sept 2014, pp. 154–159.

[133] Y.-H. Feng, T.-H. Teng, and A.-H. Tan, "Modelling situation awareness for context-aware decision support," *Expert Systems with Applications*, vol. 36, no. 1, pp. 455 – 463, 2009.

[134] M. Naderpour, J. Lu, and G. Zhang, "An intelligent situation awareness support system for safety-critical environments," *Decision Support Systems*, vol. 59, no. Supplement C, pp. 325 – 340, 2014.

[135] G. Jakobson, J. Buford, and L. Lewis, “A framework of cognitive situation modeling and recognition,” in *MILCOM 2006 - 2006 IEEE Military Communications conference*, Oct 2006, pp. 1–7.

[136] N. Baumgartner, S. Mitsch, A. Muller, W. Retschitzegger, A. Salfinger, and W. Schwinger, “A tour of BeAware - a situation awareness framework for control centers,” *Information Fusion*, vol. 20, no. Supplement C, pp. 155 – 173, 2014.

[137] A. Srivastav, Y. Wen, E. Hendrick, I. Chattopadhyay, A. Ray, and S. Phoha, “Information fusion for object situation assessment in sensor networks,” in *Proceedings of the 2011 American Control Conference*, June 2011, pp. 1274–1279.

[138] N. Brannon, G. Conrad, T. Draelos, J. Seiffertt, and D. Wunsch, “Information fusion and situation awareness using artmap and partially observable markov decision processes,” in *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, 2006, pp. 2023–2030.

[139] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, “Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps,” *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 698–713, Sep 1992.

[140] R. Pearson, M. Donnelly, J. Liu, and L. Galway, “A framework for situation awareness based upon dynamic situation modeling,” in *Ambient Assisted Living and Daily Activities: 6th International Work-Conference, IWAAL 2014, Belfast, UK, December 2-5, 2014. Proceedings.* Cham: Springer International Publishing, 2014, pp. 99–102.

[141] C. J. Matheus, M. M. Kokar, K. Baclawski, J. A. Letkowski, C. Call, M. L. Hinman, J. J. Salerno, and D. M. Boulware, “SAWA: an assistant for higher-level fusion and situation awareness,” *Proc.SPIE*, vol. 5813, pp. 5813 – 5813 – 11, 2005.

[142] C. J. Matheus, M. M. Kokar, K. Baclawski, J. J. Letkowski, C. Call, M. Hinman, J. Salerno, and D. Boulware, “Lessons

learned from developing SAWA: a situation awareness assistant," in *2005 7th International Conference on Information Fusion*, vol. 2, July 2005, p. 8.

[143] J. Salerno, M. Hinman, and D. Boulware, "Building a framework for situation awareness," in *Proc. of the Seventh Intl. Conf. on Information Fusion*, 2004, pp. 219–226.

[144] A. Ciaramella, M. G. C. A. Cimino, F. Marcelloni, and U. Straccia, "Combining fuzzy logic and semantic web to enable situation-awareness in service recommendation," in *Database and Expert Systems Applications: 21st International Conference, DEXA 2010, Bilbao, Spain, August 30 - September 3, 2010, Proceedings, Part I.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 31–45.

[145] P. R. Smart, N. R. Shadbolt, L. A. Carr, and M. C. Schraefel, "Knowledge-based information fusion for improved situational awareness," in *2005 7th International Conference on Information Fusion*, vol. 2, July 2005, pp. 8 pp.–.

[146] R. Laxhammar, "Anomaly detection for sea surveillance," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–8.

[147] M. Riveiro, G. Falkman, and T. Ziemke, "Improving maritime anomaly detection and situation awareness through interactive visualization," in *2008 11th International Conference on Information Fusion*, June 2008, pp. 1–8.

[148] M. Glandrup, "Improving situation awareness in the maritime domain," in *Situation Awareness with Systems of Systems.* New York, NY: Springer New York, 2013, pp. 21–38.

[149] P. van de Laar, J. Tretmans, and M. Borth, "Situation awareness with systems of systems," 2013.

[150] A. Van den Broek, R. Neef, P. Hanckmann, S. P. van Gosliga, and D. Van Halsema, "Improving maritime situational awareness by fusing sensor information and intelligence," in *Information*

Fusion (FUSION), 2011 Proceedings of the 14th International Conference on. IEEE, 2011, pp. 1–8.

[151] M. Gariel, A. N. Srivastava, and E. Feron, "Trajectory clustering and an application to airspace monitoring," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1511–1524, Dec 2011.

[152] S. Krishnaswamy, S. Loke, A. Rakotonirainy, O. Horovitz, and M. Gaber, "Towards situation-awareness and ubiquitous data mining for road safety: Rationale and architecture for a compelling application," in *Proceedings of Conference on Intelligent Vehicles and Road Infrastructure*, 2005.

[153] J. Rogstadius, M. Vukovic, C. A. Teixeira, V. Kostakos, E. Karapanos, and J. A. Laredo, "Crisistracker: Crowdsourced social media curation for disaster awareness," *IBM Journal of Research and Development*, vol. 57, no. 5, pp. 4:1–4:13, Sept 2013.

[154] M. A. Cameron, R. Power, B. Robinson, and J. Yin, "Emergency situation awareness from twitter for crisis management," in *Proceedings of the 21st International Conference on World Wide Web*, ser. WWW '12 Companion. New York, NY, USA: ACM, 2012, pp. 695–698.

[155] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using social media to enhance emergency situation awareness," *IEEE Intelligent Systems*, vol. 27, no. 6, pp. 52–59, Nov 2012.

[156] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111 – 127, 1997, fuzzy Sets: Where Do We Stand? Where Do We Go?

[157] W. Pedrycz, "Granular computing: an introduction," in *Joint 9th IFSA World Congress and 20th NAFIPS International Conference, 2001.*, vol. 3, July 2001, pp. 1349–1354 vol.3.

[158] Y. Yao and N. Zhong, "Granular computing," *Wiley Encyclopedia of Computer Science and Engineering*, 2008.

[159] A. Bargiela and W. Pedrycz, "The roots of granular computing," in *2006 IEEE International Conference on Granular Computing*, May 2006, pp. 806–809.

[160] G. Wilke and E. Portmann, "Granular computing as a basis of human–data interaction: a cognitive cities use case," *Granular Computing*, vol. 1, no. 3, pp. 181–197, Sep 2016.

[161] L. Zadeh, "Fuzzy sets and information granularity," *Advances in Fuzzy Set Theory and Applications, North-Holland, Amsterdam*, pp. 3–18.

[162] W. Pedrycz, "History and development of granular computing," *Encyclopedia of Life Support Systems (EOLSS). Eolss Publishers, Paris*, 2012.

[163] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: perspectives and challenges," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977–1989, 2013.

[164] Y. Yao, "The art of granular computing," in *Rough Sets and Intelligent Systems Paradigms: International Conference, RSEISP 2007, Warsaw, Poland, June 28-30, 2007. Proceedings.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 101–112.

[165] ——, "Integrative levels of granularity," in *Human-Centric Information Processing Through Granular Modelling.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 31–47.

[166] J. T. Yao, "Information granulation and granular relationships," in *2005 IEEE International Conference on Granular Computing*, vol. 1, July 2005, pp. 326–329 Vol. 1.

[167] W. Pedrycz, A. Gacek, and X. Wang, "Clustering in augmented space of granular constraints: A study in knowledge-based clustering," *Pattern Recognition Letters*, vol. 67, no. Part 2, pp. 122 – 129, 2015, granular Mining and Knowledge Discovery.

[168] W. Pedrycz, B. Park, and S. Oh, "The design of granular classifiers: A study in the synergy of interval calculus and fuzzy sets in pattern recognition," *Pattern Recognition*, vol. 41, no. 12, pp. 3720 – 3735, 2008.

[169] S. Salehi, A. Selamat, M. R. Mashinchi, and H. Fujita, "The synergistic combination of particle swarm optimization and fuzzy sets to design granular classifier," *Knowledge-Based Systems*, vol. 76, no. Supplement C, pp. 200 – 218, 2015.

[170] B. Q. Hu, "Three-way decisions space and three-way decisions," *Information Sciences*, vol. 281, no. Supplement C, pp. 21 – 52, 2014, multimedia Modeling.

[171] W. Pedrycz, R. Al-Hmouz, A. Morfeq, and A. S. Balamash, "Building granular fuzzy decision support systems," *Knowledge-Based Systems*, vol. 58, no. Supplement C, pp. 3 – 10, 2014, intelligent Decision Support Making Tools and Techniques: IDSMT.

[172] P. Grzegorzewski, "Fuzzy number approximation via shadowed sets," *Information Sciences*, vol. 225, no. Supplement C, pp. 35 – 46, 2013.

[173] M. Ye, X. Wu, X. Hu, and D. Hu, "Anonymizing classification data using rough set theory," *Knowledge-Based Systems*, vol. 43, no. Supplement C, pp. 82 – 94, 2013.

[174] T. Velmurugan, "Performance based analysis between k-means and fuzzy c-means clustering algorithms for connection oriented telecommunication data," *Applied Soft Computing Journal*, vol. 19, pp. 134–146, 2014.

[175] M. A. Sanchez, O. Castillo, J. R. Castro, and P. Melin, "Fuzzy granular gravitational clustering algorithm for multivariate data," *Information Sciences*, vol. 279, no. Supplement C, pp. 498 – 511, 2014.

[176] W. Pedrycz and W. Homenda, "Building the fundamentals of granular computing: a principle of justifiable granularity," *Applied Soft Computing*, vol. 13, no. 10, pp. 4209–4218, 2013.

[177] Y. Yao, "Three perspectives of granular computing," *Journal of Nanchang Institute of Technology*, vol. 25, no. 2, pp. 16–21, 2006.

[178] ——, "Human-inspired granular computing," *Novel developments in granular computing: applications for advanced human reasoning and soft computation*, pp. 1–15, 2010.

[179] S. Chibbaro, L. Rondoni, and A. Vulpiani, “Reductionism, emergence and levels of reality,” *Springer, Berlin*, vol. 3, no. 20, p. 17, 2014.

[180] F. Capra, *The web of life.* Audio Renaissance Tapes, 1996.

[181] E. Laszlo, “The systems view of the world a holistic vision for our time,” 1996.

[182] H. C. Brown, “Structural levels in the scientist’s world,” *The Journal of Philosophy, Psychology and Scientific Methods*, pp. 337–345, 1916.

[183] L. Floridi, “The method of levels of abstraction,” *Minds and machines*, vol. 18, no. 3, pp. 303–329, 2008.

[184] A. Bargiela and W. Pedrycz, *Granular Computing - An Introduction.* Boston, Dordrecht, London: Kluwer Academic Publishers, 2003.

[185] L. Wang, X. Yang, J. Yang, and C. Wu, “Relationships among generalized rough sets in six coverings and pure reflexive neighborhood system,” *Information Sciences*, vol. 207, no. Supplement C, pp. 66 – 78, 2012.

[186] M. G. Cimino, B. Lazzerini, F. Marcelloni, and W. Pedrycz, “Genetic interval neural networks for granular data regression,” *Information Sciences*, vol. 257, no. Supplement C, pp. 313 – 330, 2014.

[187] H. Hu, L. Pang, D. Tian, and Z. Shi, “Perception granular computing in visual haze-free task,” *Expert Systems with Applications*, vol. 41, no. 6, pp. 2729 – 2741, 2014.

[188] A. V. Nandedkar, “An interactive colour video segmentation: A granular computing approach,” in *Electrical Engineering and Intelligent Systems.* New York, NY: Springer New York, 2013, pp. 135–146.

[189] X. Yang and J. Yang, “Neighborhood system and rough set in incomplete information system,” in *Incomplete Information System and Rough Set Theory: Models and Attribute Reductions.*

Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 101–130.

[190] W. Pedrycz, W. Lu, X. Liu, W. Wang, and L. Wang, "Human-centric analysis and interpretation of time series: a perspective of granular computing," *Soft Computing*, vol. 18, no. 12, pp. 2397–2411, 2014.

[191] A. Skowron, A. Jankowski, and S. Dutta, "Interactive granular computing," *Granular Computing*, vol. 1, no. 2, pp. 95–113, Jun 2016.

[192] Y.-Y. Yao, "The rise of granular computing," *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, vol. 20, no. 3, pp. 299–308, 2008.

[193] S. Mittal, A. Aggarwal, and S. L. Maskara, "Situation recognition in sensor based environments using concept lattices," in *Proceedings of the CUBE International Information Technology Conference*, ser. CUBE '12. New York, NY, USA: ACM, 2012, pp. 579–584.

[194] M. M. Kokar, C. J. Matheus, and K. Baclawski, "Ontology-based situation awareness," *Information fusion*, vol. 10, no. 1, pp. 83–98, 2009.

[195] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[196] W. Pedrycz, G. Succi, A. Sillitti, and J. Iljazi, "Data description: A general framework of information granules," *Knowledge-Based Systems*, vol. 80, pp. 98–108, 2015.

[197] A. Balamash, W. Pedrycz, R. Al-Hmouz, and A. Morfeq, "An expansion of fuzzy information granules through successive refinements of their information content and their use to system modeling," *Expert Systems with Applications*, vol. 42, no. 6, pp. 2985–2997, 2015.

[198] W. Pedrycz and A. Gacek, "Temporal granulation and its application to signal analysis," *Information Sciences*, vol. 143, no. 1, pp. 47–71, 2002.

[199] A. Gacek, "Signal processing and time series description: A perspective of computational intelligence and granular computing," *Applied Soft Computing*, vol. 27, pp. 590–601, 2015.

[200] M. A. Sanchez, O. Castillo, and J. R. Castro, "Information granule formation via the concept of uncertainty-based information with interval type-2 fuzzy sets representation and takagi–sugeno–kang consequents optimized with cuckoo search," *Applied Soft Computing*, vol. 27, pp. 602–609, 2015.

[201] D. Sánchez, P. Melin, and O. Castillo, "Optimization of modular granular neural networks using a hierarchical genetic algorithm based on the database complexity applied to human recognition," *Information Sciences*, vol. 309, pp. 73–101, 2015.

[202] S. Meher and D. Kumar, "Ensemble of adaptive rule-based granular neural network classifiers for multispectral remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, no. 99, pp. 1–10, 2015.

[203] Y.-Q. Zhang, M. D. Fraser, R. Gagliano, A. Kandel *et al.*, "Granular neural networks for numerical-linguistic data fusion and knowledge discovery," *Neural Networks, IEEE Transactions on*, vol. 11, no. 3, pp. 658–667, 2000.

[204] J. F. Peters, S. Ramanna, A. Skowron, J. Stepaniuk, and Z. Suraj, "Sensor fusion: A rough granular approach," in *9th IFSA World Congress and 20th NAFIPS International Conference, 2001*, vol. 3. IEEE, 2001, pp. 1367–1371.

[205] W. Haijun and C. Yimin, "Sensor data fusion using rough set for mobile robots system," in *Mechatronic and Embedded Systems and Applications, Proceedings of the 2nd IEEE/ASME International Conference on.* IEEE, 2006, pp. 1–5.

[206] T. Guan and B. Feng, "Rough fuzzy integrals for information fusion and classification," in *Rough Sets and Current Trends in Computing.* Springer, 2004, pp. 362–367.

[207] F. Jiang, Y. Sui, and C. Cao, "Outlier detection using rough set theory," in *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing.* Springer, 2005, pp. 79–87.

[208] T. T. Nyuyen, "Outlier and exception analysis in rough sets and granular computing," *Handbook of Granular Computing*, pp. 823–834, 2008.

[209] F. Shaari, A. A. Bakar, and A. R. Hamdan, "Outlier detection based on rough sets theory," *Intelligent Data Analysis*, vol. 13, no. 2, pp. 191–206, 2009.

[210] Y. Chen, D. Miao, and H. Zhang, "Neighborhood outlier detection," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8745–8749, 2010.

[211] F. Jiang and Y.-M. Chen, "Outlier detection based on granular computing and rough set theory," *Applied Intelligence*, vol. 42, no. 2, pp. 303–322, 2015.

[212] A. Albanese, S. K. Pal, and A. Petrosino, "Rough sets, kernel set, and spatiotemporal outlier detection," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 26, no. 1, pp. 194–207, 2014.

[213] P. K. Dutta, O. Mishra, and M. Naskar, "Improving situational awareness for precursory data classification using attribute rough set reduction approach," *International Journal of Information Technology and Computer Science (IJITCS)*, vol. 5, no. 12, p. 47, 2013.

[214] X. Jia, L. Shang, B. Zhou, and Y. Yao, "Generalized attribute reduction in rough set theory," *Knowledge-Based Systems*, 2015.

[215] R. R. Yager, "A framework for multi-source data fusion," *Information Sciences*, vol. 163, no. 1, pp. 175–200, 2004.

[216] E. Herrera-Viedma, F. J. Cabrerizo, J. Kacprzyk, and W. Pedrycz, "A review of soft consensus models in a fuzzy environment," *Information Fusion*, vol. 17, pp. 4–13, 2014.

[217] Y. Yao, "Interpreting concept learning in cognitive informatics and granular computing," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 4, pp. 855–866, 2009.

[218] J. Li, C. Mei, W. Xu, and Y. Qian, “Concept learning via granular computing: A cognitive viewpoint,” *Information Sciences*, vol. 298, pp. 447–467, 2015.

[219] W. Homenda and W. Pedrycz, “Linguistic approach to granular cognitive maps,” in *Intelligent Systems’2014*, ser. Advances in Intelligent Systems and Computing. Springer International Publishing, 2015, vol. 322, pp. 205–216.

[220] P. K. Singh, C. A. Kumar, and J. Li, “Knowledge representation using interval-valued fuzzy formal concept lattice,” *Soft Computing*, pp. 1–18, 2015.

[221] W.-Z. Wu, Y. Leung, and J.-S. Mi, “Granular computing and knowledge reduction in formal contexts,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 10, pp. 1461–1474, 2009.

[222] A. Jankowski, A. Skowron, and R. Swiniarski, “Interactive rough-granular computing in wisdom technology,” in *Active Media Technology.* Springer, 2013, pp. 1–13.

[223] A. Skowron and A. Jankowski, “Interactive computations: toward risk management in interactive intelligent systems,” *Natural Computing*, pp. 1–12, 2015.

[224] R. D. Fricker, *Introduction to Statistical Methods for Biosurveillance: With an Emphasis on Syndromic Surveillance.* Cambridge University Press, 2013.

[225] R. Al-Hmouz, W. Pedrycz, and A. Balamash, “Description and prediction of time series: A general framework of granular computing,” *Expert Systems with Applications*, vol. 42, no. 10, pp. 4830–4839, 2015.

[226] W. Wang, W. Pedrycz, and X. Liu, “Time series long-term forecasting model based on information granules and fuzzy clustering,” *Engineering Applications of Artificial Intelligence*, vol. 41, pp. 17–24, 2015.

[227] W. Lu, J. Yang, and X. Liu, “Numerical prediction of time series based on fcms with information granules,” *International Journal of Computers Communications & Control*, vol. 9, no. 3, pp. 313–324, 2014.

[228] M. Endsley and D. Garland, *Situation Awareness Analysis and Measurement*. CRC Press, 2000.

[229] R. W. Pew, “The state of situation awareness measurement: Heading toward the next century,” in *Situation Awareness Analysis and Measurement*, M. Endsley and D. Garland, Eds. CRC Press, 2000, pp. 33–47.

[230] R. M. Taylor, “Situational awareness rating technique (SART): The development of a tool for aircrew systems design,” in *Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478*. Seuilly-sur Seine: NATO AGARD, 1989.

[231] M. R. Endsley, “Direct measurement of situation awareness: Validity and use of SAGAT,” in *Situation Awareness Analysis and Measurement*, M. Endsley and D. Garland, Eds. CRC Press, 2000, pp. 147–173.

[232] A. Boukerche and S. Samarah, “A novel algorithm for mining association rules in wireless ad hoc sensor networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, no. 7, pp. 865–877, July 2008.

[233] E. C.-H. Ngai and P. Gunningberg, “Quality-of-information-aware data collection for mobile sensor networks,” *Pervasive and Mobile Computing*, vol. 11, pp. 203 – 215, 2014.

[234] Z. Qin, Q. Han, S. Mehrotra, and N. Venkatasubramanian, “Quality-aware sensor data management,” in *The Art of Wireless Sensor Networks: Volume 1: Fundamentals*, 2014, pp. 429–464.

[235] V. Sachidananda, A. Khelil, and N. Suri, “Quality of information in wireless sensor networks: A survey,” in *Proc. of the 15th International Conference on Information Quality*, 2010, pp. 193–207.

[236] T.-P. Hong and C.-W. Wu, "Mining rules from an incomplete dataset with a high missing rate," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3931 – 3936, 2011.

[237] W. Wang, S. De, R. Toenjes, E. Reetz, and K. Moessner, "A comprehensive ontology for knowledge representation in the internet of things," in *2012 IEEE 11th International Conference onTrust, Security and Privacy in Computing and Communications (TrustCom)*, 2012, pp. 1793–1798.

[238] M. Compton, P. Barnaghi, L. Bermudez, R. GarcÃa-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. D. Kelsey, D. L. Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, and K. Taylor, "The SSN ontology of the W3C semantic sensor network incubator group," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 17, pp. 25 – 32, 2012.

[239] D. Martin, M. Burstein, D. McDermott, S. McIlraith, M. Paolucci, K. Sycara, D. L. McGuinness, E. Sirin, and N. Srinivasan, "Bringing semantics to web services with OWL-S," *World Wide Web*, vol. 10, no. 3, pp. 243–277, 2007.

[240] B. Pernici and S. H. Siadat, "A fuzzy service adaptation based on qos satisfaction," in *Advanced Information Systems Engineering: 23rd International Conference, CAiSE 2011, London, UK, 2011.*, pp. 48–61.

[241] Z. Xiao and G. Wei, "Application interval-valued intuitionistic fuzzy set to select supplier," in *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 3, Oct 2008, pp. 351–355.

[242] G. P. Timms, P. A. de Souza, L. Reznik, and D. V. Smith, "Automated data quality assessment of marine sensors," *Sensors*, vol. 11, no. 10, p. 9589, 2011.

[243] N. Jiang and L. Gruenwald, "Estimating missing data in data streams," in *Proceedings Advances in Databases: Concepts,*

Systems and Applications: 12th International Conference on Database Systems for Advanced Applications, DASFAA 2007, April 9-12, 2007. Bangkok, Thailand: Springer Berlin Heidelberg, pp. 981–987.

[244] J. L. Schafer and J. W. Graham, "Missing data: our view of the state of the art." *Psychological methods*, vol. 7, no. 2, p. 147, 2002.

[245] B. M. Patil, R. C. Joshi, and D. Toshniwal, "Missing value imputation based on k-mean clustering with weighted distance," in *Contemporary Computing: Third International Conference, IC3 2010, Noida, India, August 9-11, 2010. Proceedings, Part I.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 600–609.

[246] G. E. Batista and M. C. Monard, "A study of k-nearest neighbour as an imputation method," *Soft Computing Systems: Design, Management and Applications*, p. 251–260, 2002.

[247] P. D. Allison, *Missing data.* Sage Thousand Oaks, CA, 2012.

[248] P. Baraldi, F. D. Maio, D. Genini, and E. Zio, "Reconstruction of missing data in multidimensional time series by fuzzy similarity," *Applied Soft Computing*, vol. 26, pp. 1 – 9, 2015.

[249] T.-P. Hong, L.-H. Tseng, and B.-C. Chien, "Mining from incomplete quantitative data by fuzzy rough sets," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2644 – 2653, 2010.

[250] E.-L. Silva-Ramirez, R. Pino-Mejias, and M. Lopez-Coello, "Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbours for monotone patterns," *Applied Soft Computing*, vol. 29, pp. 65 – 74, 2015.

[251] C.-H. Wu, C.-H. Wun, and H.-J. Chou, "Using association rules for completing missing data," in *Hybrid Intelligent Systems, 2004. HIS '04. Fourth International Conference on*, Dec 2004, pp. 236–241.

[252] J.-J. Shen, C.-C. Chang, and Y.-C. Li, "Combined association rules for dealing with missing values," *Journal of Information Science*, vol. 33, no. 4, pp. 468–480, 2007.

[253] Z. Gao, W. Cheng, X. Qiu, and L. Meng, "A missing sensor data estimation algorithm based on temporal and spatial correlation," *International Journal of Distributed Sensor Networks*, vol. 11, no. 10, p. 435391, 2015.

[254] A. Ragel and B. Cremilleux, "MVC - a preprocessing method to deal with missing values," *Knowledge-Based Systems*, vol. 12, no. 5, pp. 285–291, 1999.

[255] Smile. (2016) Statistical machine intelligence and learning engine (smile). [Online]. Available: http://haifengl.github.io/smile/

[256] G. Jakobson, J. F. Buford, and L. Lewis, "Situation management [guest editorial]," *IEEE Communications Magazine*, vol. 48, no. 3, pp. 110–111, 2010.

[257] L. O. Bonino da Silva Santos and et al., "GSO: Designing a well-founded service ontology to support dynamic service discovery and composition," in *13th Enterprise Distributed Object Computing Conf.*, September 2009, pp. 35–44.

[258] C. J. Matheus, M. M. Kokar, and K. Baclawski, "A core ontology for situation awareness," in *Sixth International Conference of Information Fusion, 2003. Proceedings of the*, vol. 1, July 2003, pp. 545–552.

[259] A. Khatami, P. Huibers, and J. J. Roessingh, "Architecture for goal-driven behavior of virtual opponents in fighter pilot combat training," in *Proceedings of the 22nd Annual Conference on Behavior Representation in Modeling and Simulation*, 2013.

[260] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press Cambridge, 1998, vol. 1, no. 1.

[261] R. S. Sutton, "Generalization in reinforcement learning: Successful examples using sparse coarse coding," *Advances in neural information processing systems*, pp. 1038–1044, 1996.

[262] R. Azevedo and J. Cromley, "Does training on self-regulated learning facilitate students' learning with hypermedia?" *Journal of Educational Psychology*, vol. 96, no. 3, pp. 523–535, 2004.

[263] S. Nagulendra and J. Vassileva, "Minimizing social data overload through interest-based stream filtering in a P2P social network," in *International Conference on Social Computing, SocialCom 2013, Washington, DC, USA, 8-14 September, 2013*, 2013, pp. 878–881.

[264] M. R. Endsley, "Direct measurement of situation awareness:validity and use of SAGAT," in *Situation awareness analysis and measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2000.

[265] H. Billhardt, A. Fernandez, L. Lemus, M. Lujak, N. Osman, S. Ossowski, and C. Sierra, "Dynamic coordination in fleet management systems: Toward smart cyber fleets," *Intelligent Systems, IEEE*, vol. 29, no. 3, pp. 70–76, 2014.

[266] A. Cartenì, G. Cantarella, and S. DELUCA, "A simulation model for a container terminal," *Proceedings of ETC 2005, transport policy and operations-freight and logistics-ports*, 2005.

[267] M. Antonelli, P. Ducange, B. Lazzerini, and F. Marcelloni, "Multi-objective evolutionary design of granular rule-based classifiers," *Granular Computing*, vol. 1, no. 1, pp. 37–58, Mar 2016.

[268] W. Pedrycz, "Evolvable fuzzy systems: some insights and challenges," *Evolving Systems*, vol. 1, no. 2, pp. 73–82, Sep 2010.

[269] D. Leite, P. Costa, and F. Gomide, "Interval approach for evolving granular system modeling," in *Learning in Non-Stationary Environments: Methods and Applications*. New York, NY: Springer New York, 2012, pp. 271–300.

[270] Y. Yao, "Granular computing using neighborhood systems," in *Advances in Soft Computing*. Springer, 1999, pp. 539–553.

[271] J. Liang, "Uncertainty and feature selection in rough set theory," in *Rough Sets and Knowledge Technology: 6th International*

Conference, RSKT 2011, Banff, Canada, October 9-12, 2011. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 8–15.

[272] Y. Qian, Y. Li, J. Liang, G. Lin, and C. Dang, "Fuzzy granular structure distance," *Fuzzy Systems, IEEE Transactions on*, vol. 23, no. 6, pp. 2245–2259, 2015.

[273] R. L. Newman, "Scenarios for rare event simulation and flight testing," *Crew Systems TR-02-07A, Monterey Technologies Inc*, 2002.

[274] B. Liu, W. Hsu, L.-F. Mun, and H.-Y. Lee, "Finding interesting patterns using user expectations," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 11, no. 6, pp. 817–832, 1999.

[275] B. Kosko, *Fuzzy Engineering.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.

[276] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1, pp. 1–6, 1998.

[277] S. A. Mingoti and J. O. Lima, "Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms," *European Journal of Operational Research*, vol. 174, no. 3, pp. 1742–1759, 2006.

[278] Z. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[279] T.-P. Hong, L.-H. Tseng, and B.-C. Chien, "Mining from incomplete quantitative data by fuzzy rough sets," *Expert Systems with Applications*, vol. 37, no. 3, pp. 2644 – 2653, 2010.

[280] A. Hassanien, A. Abraham, J. Peters, G. Schaefer, and C. Henry, "Rough sets and near sets in medical imaging: A review," *IEEE Trans. on Information Technology in Biomedicine*, vol. 13, no. 6, 2009.

[281] D. Van Nguyen, K. Yamada, and M. Unehara, "Extended tolerance relation to define a new rough set model in incomplete

information systems," *Adv. Fuzzy Sys.*, vol. 2013, pp. 9:9–9:9, Jan. 2013.

[282] Y. Yao, "Probabilistic rough set approximations," *International Journal of Approximate Reasoning*, vol. 49, no. 2, pp. 255 – 271, 2008.

[283] S. Li and T. Li, "Incremental update of approximations in dominance-based rough sets approach under the variation of attribute values," *Information Sciences*, vol. 294, pp. 348–361, 2015.