Individualized Situation Recognition using Approximate Case-Based Reasoning

Von der Fakultät für Ingenieurwissenschaften,
Abteilung Maschinenbau und Verfahrenstechnik
der
Universität Duisburg-Essen
zur Erlangung des akademischen Grades
einer
Doktorin der Ingenieurwissenschaften
Dr.-Ing.
genehmigte Dissertation

von

Arezoo Sarkheyli-Hägele
aus
Mashhad, Iran

Gutachter: Univ.-Prof. Dr.-Ing. Dirk Söffker, Universität Duisburg-Essen
Prof. Dr. Stefan Kopp, Universität Bielefeld

Tag der mündlichen Prüfung: 20.11.2017

Acknowledgements

The work presented herein is nonetheless at least as much a product of four years doctoral research at the Chair Dynamics and Control (SRS) at the University of Duisburg-Essen.

First and foremost I would like to express my special appreciation and thanks to my supervisor, Prof. Dirk Söffker, for the guidance and great support he has given me during over the years. It has been a big honor to work with him and learn from him. I appreciate all his contributions of time, ideas, and supports to enhance my experience. His advices on both research as well as on my career and life have been invaluable.

I want to thank German Academic Exchange Service (Deutscher Akademischer Austausch Dienst - DAAD) and its staff who make this research possible by providing financial support. This work could not been accomplished without the DAAD support.

I would like to thank Prof. Wolfram Luther for great discussions and valuable comments during my research as well as his recommendations. I would also like to thank Prof. Stephan Kopp for giving numerous helpful suggestions and great questions to improve this research.

I wish to express my warmest gratitude to all my friends, old and present colleagues at the Chair of Dynamics and Control (SRS) whose comments and questions as well as their personal and academic supports have left a mark on this work. My special thanks go to my colleague Jiao Wang for her support and great help with the experiments related to this work. I particularly like to thank the friends who assisted me in experimental data acquisition.

Finally, I owe everything to my family whose their academically and emotionally supports through all the difficulties in this research journey are indescribable. I would like to extend my sincere gratitude to my husband who is also my best friend and colleague, Georg Hägele for the suggestions, insightful questions, criticisms, and all his supports throughout this work. Thank you for encouraging me throughout this journey.

Duisburg, November 2017 Arezoo Sarkheyli-Hägele

Kurzfassung

Situationserkennung ist sowohl ein wichtiger Aspekt der menschlichen Wahrnehmung wie auch bei der Überwachung der Entscheidungsfindung eines menschlichen Nutzers (human operator) in ungeplanten, ungenauen und unsicheren Umgebungen. Die Situationserkennung ist ein Prozess zur Identifizierung der aktuellen Situation, die aus Änderungen innerhalb der Umgebung resultiert. Für verschiedene Anwendungen wurden verschiedene Situationserkennungsansätze erfolgreich entwickelt. Die Individualisierung der Situationserkennung wurde in der Literatur bisher wenig beachtet. Der Situationserkennungsprozess konnte für die Unterstützung der menschlichen Nutzer durch das Lernen und Berücksichtigung individueller Verhaltensweisen, Präferenzen und Prioritäten verbessert werden.

Ereignisdiskrete Situationen, die aus einer Folge von Aktionen resultieren, können individuelle Verhaltensweisen menschlicher Nutzer repräsentieren. In dieser Arbeit werden Repräsentation und Identifizierung von ereignisdiskreten Situationen betrachtet.

In dieser Arbeit wird ein neues Framework für eine individualisierte Situationserkennung vorgeschlagen, indem neuartige Wissensrepräsentationen und Schlussfolgerungsansätze angewendet werden. Die wichtigsten Herausforderungen die durch die eingeführten Frameworks für die individualisierte Situationserkennung zu lössen sind, sind die Modellierung und Repräsentation von Wissen sowie das Lernen auf neuen, unbekannten Situationen. Diese Herausforderungen können basierend wie folgt angegeben werden: Wie können die ereignisdiskreten Situationen modelliert und in einen Wissensspeicher übertragen werden? Wie kann das gespeicherte Wissen für die weitere Situationserkennung verwendet werden?

In dieser Arbeit wird der Case-Based Reasoning (CBR)-Ansatz verwendet, um eine individualisierte Situationserkennung zur Unterstützung menschlicher Nutzer zu realisieren. Der klassische CBR-Ansatz wird um einen neuen Lernprozessansatz erweitert, sodass bekannte Situationen erkannt werden können und neues Wissen aus unbekannten Situationen generiert werden kann. Um die erwähnten Herausforderungen zu bewältigen und die Situationerkennung zu realisieren, wird der klassische CBR-Ansatz um den Situation-Operator Modeling (SOM)-Ansatz und Fuzzy-Logik (FL)-Ansatz erweitert. Der SOM-Ansatz ermöglicht eine wissensorientierte Modellbildung und wird zur strukturierten Beschreibung der Beziehungen der ereignisdiskreten Situationen innerhalb dynamischer Umgebungen verwendet. Der präsentierte SOM-Ansatz wird innerhalb des Lernprozesses dazu verwendet, Sequenzen von Situationen und Aktionen einzelnen Situationsmustern zuzuordnen. Der FL-Ansatz strukturiert das SOM-basierte Wissen für approximate reasoning. Zusätzliche Prozesse werden zur Unterstützung des Onlinelernens, der Datenreduktion, und Wissensindexierung inerhalb des vorschlagenen fuzzy SOM-basierte CBR-Ansatzes verwendet.

Die Funktionsweise des vorgestellten Frameworks wird am Beispiel der Realisierung einer individualisierten Spurwechsel-Situationserkennung zur Unterstützung menschlicher Fahrer verdeutlicht. Hierbei ist das Ziel, eine passende Fahrsituation zum Spurwechsel zu erkennen. Das Framework wird unter Verwendung von Testdaten evaluiert. Die Datenerfassung erfolgt auf Basis eines Fahrsimulators und unterschiedlicher Testfahrer. Hierauf aufbauend wird der Nachweis einer erfolgreichen Verwendung des vorgestellten Anzatzes zur individualisierten Situationserkennung erbracht.

Die Ergebnisse zeigen, dass unter Verweldung des vorgestellten Ansatzes eine erfolgreiche Fahrsituationerkennung realisiert werden kann, im Sinne von Genauigkeit, Erkennungsrate, negativer Alarmrate und Erkennungszeit. Abschliessend wird gezeigt, dass eine Individualisierung die Genauigkeit der Situationserkennung signifikant verbessert werden kann.

Abstract

Situation recognition is a significant part of humans perception as well as in the process of supervising human operators decision making in unplanned, imprecise, and uncertain environments. It is a process for identification of actual situation as the result of the occurring events within the environment. With outstanding performance, different situation recognition approaches for various applications have been developed. However, far too little attention has been paid to individualization of situation recognition. Situation recognition process could be individualized for supervision of human operators by learning and considering exclusive behaviors, preferences, and priorities of individual human operators.

The event-discrete situations which are generated with a sequence of triggered actions could express individual behaviors of human operators. Accordingly, representation and identification of event-discrete situations are considered in this contribution.

The purpose of this thesis is to propose a new framework for individualized situation recognition by applying novel knowledge representation and reasoning approaches. The most major challenges to be solved through the proposed framework for individualized situation recognition are modeling and representation of experienced knowledge as well as learning the new unknown situations. Those challenges could be stated as two questions as follows: How to model and represent the event-discrete situations to a knowledge base? How to reuse the knowledge for further situation recognition?

In this work, Case-Based Reasoning (CBR) approach is applied to realize individualized situation recognition for supervision of human operators. The classical CBR is improved with a new learning process to recognize known occurring situations and generate new knowledge from unknown occurring situations. To deal with the noted challenges and realize the situation recognition, the classical CBR is also improved with application of a knowledge representation approach based on Situation-Operator Modeling (SOM) and fuzzy logic (FL). This work details the proposed CBR approach as a part of approximate reasoning. An integrated knowledge representation approach based on SOM and FL is introduced for representation of knowledge in the CBR. The SOM approach models the knowledge and describes the relations between discrete-event situations in a dynamic environment. The presented SOM approach supports the learning process by defining a sequence of situations and actions for each situation pattern. The FL approach structures the knowledge modeled by SOM for approximate knowledge inference. Additional processes need to be carried out in the proposed fuzzy SOM-based CBR to support online learning, data reduction, and knowledge indexing.

The presented framework is applied for realization of an individualized lane-change situation recognition to supervise human drivers. The goal is to recognize the suitable driving situations for changing the lane for individual human drivers. The

framework is evaluated using various data acquired by a driving simulator. This evaluation is done using different test drivers to highlight the effectiveness of the proposed approach for individualized situation recognition.

The results demonstrate that the proposed framework can realize a successful driving situation recognition in terms of accuracy, detection rate, false alarm rate, and recognition elapsed time. It is shown that individualized situation recognition can significantly improve the recognition accuracy.

Contents

List of Figures

List of Tables

Nomenclature

Symbols

DB_{IX}	Database of information about indices
DB_{FDP}	Database of membership functions design parameters
$CaseBase_{fuzzy}$	Fuzzy case base
$CaseBase_{indexed}$	Indexed case base
S_i	Actual situation
S_{i+1}	Following situation
S_{i-1}	Previously visited situation
O	Active action/ operator
$x_f(x)$	Membership vector of corresponding fuzzy value for a crisp value x
$\mu_{wLi}(x)$	Membership functions
wLi	Linguistic values
$thresh$	Threshold for comparing the density of the neighbors in clustered data C
W_s	Vector of features impact factors
$N(x)$	Neighborhood membership function of data point x
R_B	Equivalent relation for the subset of characteristics B
T_F	Fuzzification time
T_S	Retrieve time
T_U	Reuse time
T_I	Identification time
T_L	Labeling time
T_T	Case reduction and storage time
T_{FS}	Feature selection time
T_{IX}	Feature indexing time
T_{MFG}	Membership functions generation time
$N_{CBRQueries}$	Number of queries executed on case base
$N_{CBRUpdates}$	Number of updates of case base
$N_{CBRQueries}$	Number of updates of indexing database
$N_{CBRQueries}$	Number of updates of FDP database
P_1	Approximate start point of a lane-change maneuver
P_2	Approximate middle point of a lane-change maneuver
P_3	Exact lane-change point
P_4	Approximate end point of a lane-change maneuver
V_{a1}	Variable of ego-vehicle velocity
V_{a2}	Variable of distance to other vehicles
V_{a3}	Variable of other vehicles velocity
V_{a4}	Variable of time to collision

V_{a5}	Variable of vehicle angle
V_{a6}	Variable of steering wheel angle

Abbreviations

CBR	Case-Based Reasoning
RBR	Rule-Based Reasoning
SOM	Sitaution-Operator Modeling
FL	Fuzzy logic
HMM	Hidden Markov Model
BN	Bayesian network
FS	Fuzzy set
FCM	Fuzzy cognitive map
k-NN	k-nearest neighbor
RST	Rough set theory
FRS	Fuzzy rough set
ANN	Artificial neural network
FDT	Fuzzy decision tree
GA	Genetic algorithm
PSO	Particle swarm optimization
ACO	Ant colony optimization
HFLTS	Hesitant fuzzy linguistic term sets
NMGRS	Neighborhood-based multigranulation rough set
BDD-FCMs	Belief-degree-distributed fuzzy cognitive maps
MAS	Multi-agent system
MF	Membership function
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
FJP	Fuzzy Joint Points
NRFJP	Noise-Robust FJP
FN-DBSCAN	Fuzzy Neighborhood Density-Based Spatial Clustering of Applications with Noise
RNG	Relative neighborhood graph
MS	Mean Selection
HS	Half Selection
ST	Selection by Threshold
QR	Quick Reduction
TP	True Positive
FN	False Negative
TN	True Negative
FP	False Positive
DR	Detection rate
FA	False alarm

ACC	Accuracy
TTC	Time to collision
LCL	Lane-change to left
LCR	Lane-change to right
LK	Lane-keeping

1 Introduction

Situation recognition for a system which is known as understanding the significance of the system state and its related environment plays a key role in situation awareness. This understanding is realized by assigning the meanings to the perceived data [YDM12].

Accordingly, situation recognition could be introduced basically as a process used to support actions (action taking) and forthcoming predictions for the actual situation as well as complex decision-making in cognitive systems.

Situation recognition could be considered as an important task of cognitive systems. It is possible through reasoning in unplanned, imprecise, and uncertain situations. However, a complete description of a situation may not be possible since too much information are required to be analyzed. But, the significance of the situations with respect to the system operator goals should be recognized.

With the growing interest in using assistance systems for supervision of human operators, development of an advanced situation awareness got special attention at different application levels. In a dynamic environment, different factors such as the existence of blind-spot areas or high levels of acute stress may decrease awareness and actual understanding of the environment [PTH+16]. Situation awareness as a critical aspect of human decision-making [EG00] could improve the decisions and actions of individual human operators.

In the past several years, significant steps have been made in the improvement of situation awareness in dynamic environments. Being aware of status, characteristics, and dynamics of relevant elements of the dynamic environment and understanding the significance of those elements by considering operator goals enhance situation awareness [End95].

Moreover, real-time situation recognition plays an important role in situation awareness of human operators. Since humans usually characterize and interpret the situations according to their individual experiences and priorities [SJ16], it may be argued that consideration of exclusive behaviors of human operators improves individual situation awareness.

1.1 Towards individualized situation recognition

1.1.1 Definition of situation

A situation is defined in different ways. According to Endsley [end], a situation is the result of "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in

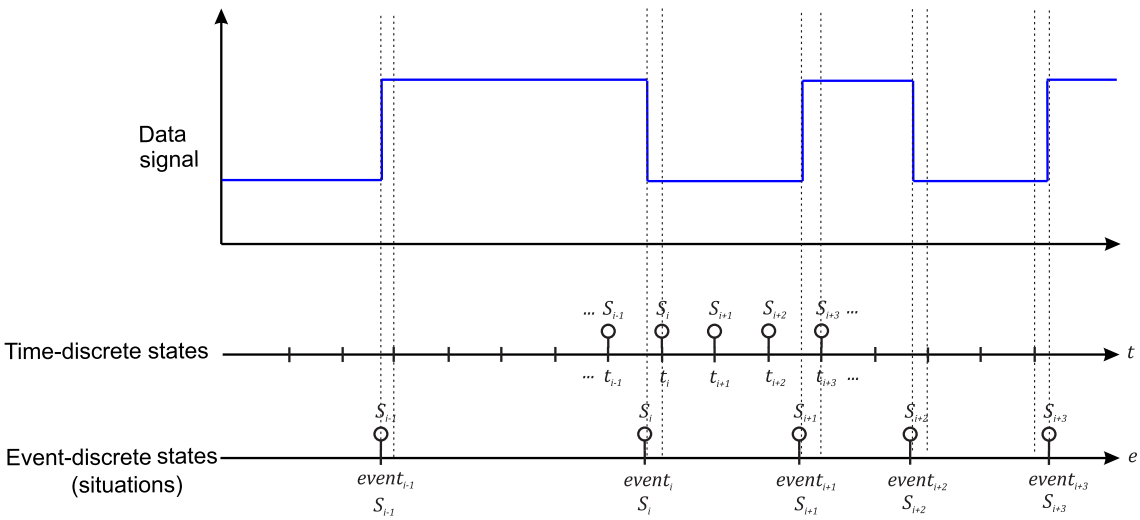

Figure 1.1: Different types of situations [SHSon]

the near future". Mooray and Sheridan [MS05] note that situation "is a shorthand description for keeping track of what is going on around you in a complex, dynamic environment" [MH87]. According to McCarthy et al. [MH87], "a situation is a finite sequence of actions". Moreover, Dousson et al. [DGG93] give the meaning of "set of event patterns and a set of constraints".

However, a situation could be defined by perceiving the status, attributes, and dynamics of relevant elements in the environment [End16]. According to [JBL06], the dynamic component defining the behaviors of an agent is based on the concepts of situations, events, actions, and time.

According to [Söf01], an event-discrete situation is generated when an event occurs. The difference between time-discrete states and event-discrete situations is shown graphically in Fig. 1.1. Here, events occur with any changes in the elements or relations in the situations. Trigger actions as being possibly the cause of an event could be defined as the functions for generating an upcoming situation.

Trigger actions could also be divided into active or passive actions [Söf01]. Active actions refer to the actions which are observable and adjustable by human operators while the passive actions may be unknown and performed by other agents in the environments.

In [Söf01], a situation S_i is defined using a set of characteristics C defining that actual state and relationships between the characteristics R.

$$\begin{aligned} S_i &= \{C, R\} \,, \\ C &= \{c_1, c_2, ..., c_n\} \,, \\ R &= \{r_1, r_2, ..., r_m\} \,, \end{aligned} \tag{1.1}$$

where n and m indicate alternatively the number of characteristics and relations defining the situation.

In addition, different relationships such as generalization (if a situation is more general than another situation), composition (if a situation is composed of other situations), dependence (if a situation occurs if and only if another situation occurs), contradiction (if two situations could not occur simultaneously), or temporal sequence (if a situation may occur after or before another situation) could exist between two situations [YDM12].

1.1.2 Problem of situation recognition

The term situation recognition to which this work is devoted is applied based on the definitions proposed by [NR03] and [Fis14]. According to [NR03], it is defined as the recognition of the situations affected by a sequence of events or actions performed by operators. Moreover, situation recognition is addressed by [Fis14] as identification of the situations possibly leading to a specific situation pattern.

According to [Dev06], the world consists not just of objects, or of objects, properties, and relations, but of objects having properties and standing in relations to one another. Consequently, a situation is describing the actual state of the world including the objects, their features, and the relations.

The essential problem of situation recognition for an agent here is understanding its actual situation (internal state of the agent and its related environment including the state of entities) at an instant of time by considering a sequence of the performed actions, occurring or upcoming situations.

A situation could be identified in three phases: (1) finding the similarity of the observed features in situations, (2) finding the relation between the features, (3) detecting the relation between the occurring situations [RC12]. Here, situation recognition is a pattern recognition problem to identify the actual situation by comparison with the defined situation patterns. Features and the relationships between the features need to be associated with the values that can change over time [BKM+03].

According to [SJ16], "the unique nature of situation recognition problem is reflected in the spatiotemporal grounding of all data as well as features defined and obtaining actionable insights from observed spatiotemporal data".

Based on the formulated situation presented by [Söf01], the situations $S_1 = \{C_1, R_1\}$ and $S_2 = \{C_2, R_2\}$ are equal if

$$\begin{aligned} C_1 &= C_2 \,, \\ R_1 &= R_2 \,. \end{aligned} \tag{1.2}$$

The set of characteristics C_1 and C_2 describing the situations must be the same. The set of relations R_1 and R_2, must relate the corresponding situation characteristics C_1 and C_2.

The situation recognition devoted to this work for a system receives a stream of time-stamped status of ego-system and its related environment as input. It performs a situation when an event occurs. It returns identified situation patterns as the output by generating deducted situations according to the trigger actions.

1.1.3 Individualized definition of situations

A situation may not always be interpreted similarly for the human operators with different personal preferences and priorities. In addition, the current environmental situation may dynamically be captured based on different emotional states of human operators [WB12]. This difference may be in terms of characteristics importance in the definition of a situation, and linguistic modifiers applied for description of the situation. However, a variety of human operator goals and experiences cause these differences.

Linguistics terms could be defined for human operators based on their acquired experiences and knowledge about the situations. A cognitive system should be able to supervise and interact with human operators identically using their acquisitive linguistics terms. Accordingly, the situation recognition process should support the situation awareness by consideration of individual human operators and individualized definition of situations.

1.2 Problem formulation

Situation recognition can be considered as a pattern matching problem where the patterns state the situation types. Here, the situations could be classified into different situation patterns. At least three important factors need to be considered for addressing a situation recognition problem as follows

- A suitable knowledge representation approach should be proposed for modeling the situations. The represented knowledge should be understandable by human decision makers and be suitable for online learning by facilitating knowledge adaption.
- Situation patterns (types) should be defined for classification of situations. These patterns are usually characterized according to human operator goals.
- It is vital to have a comprehensive and efficient approach to perform a complete process of situation recognition and learning.

It is hypothesized that the main difficulty is dealing with a complex and continuous environment specified through event-discrete, imprecise, and uncertain knowledge.

Although different researchers have focused on recognition of the situations, a real-time recognition is still sluggish. On the one hand, the notion of situations is still vague, and there is a lack of knowledge representation approach for modeling the event-discrete knowledge about dynamic environment [SJ16]. On the other hand, less attention has been paid for individualization of situation recognition.

In recent years, the problem of situation recognition has been addressed using different methods such as deterministic, probabilistic, and approximate reasoning.

The deterministic reasoning could be realized using rule-based systems, [WGB08], ontology, and Petri net [LBRR07]. This is a desirable reasoning for pattern recognition when training data does not exist. However, the definition of reasoning rules is a challenge in this type of reasoning.

The probabilistic reasoning is realized by applying conditional probability to express the relationships between characteristics of situations and representation of knowledge. From this type of reasoning, Hidden Markov Model (HMM) and Bayesian network (BN) have been widely applied for dynamic behavior modeling and activities recognition [CX04, FB12].

The approximate reasoning could be realized using a rule-like representation of knowledge to describe the relationship between characteristics of situations by application of fuzzy logic for knowledge inference [SZ14, Bal79].

Approximate and probabilistic methods are uncertainty reasoning methods and comparable with human type reasoning [Yag09]. They are the most common methods dealing with uncertain and imprecise data exhibiting various degrees of dynamism since they allow for uncertainties in knowledge representation and reasoning. According to a comparison test presented in [Yag09] using a general framework for reasoning with uncertainty based on DempsterShafer theory [Sha92], there is an underlying unity in the inferred results between these two methods.

The main difficulties of using probabilistic logics are discretization of continuous variables, providing model structure, expressing the knowledge in probability distributions, and computational complexity [PU15, Uus07]. In addition, disability of fuzzy logic in learning as well as specification of fuzzy rules without using expert knowledge are the main challenges in the application of fuzzy logic-based approximate reasoning which handles a large number of noisy sensor data [YDM12]. However, fuzzy logic as a Rule-Based Reasoning is a knowledge-based approach closely related to expert systems. According to [Ber11], Rule-Based Reasoning (RBR) may be integrated with Case-Based Reasoning (CBR) when the rules are deficient or hardly definable in some way. The CBR approach has the abilities to learn from new knowledge and situation recognition using previous knowledge.

Therefore, by considering advantages and disadvantages of the proposed approaches, the following general problem statement is formulated.

> **Problem statement:** *Is an integration of CBR with approximate reasoning approach viable for recognition of event-discrete situations in dynamic environments?*

According to [Dah11], a viable solution for situation recognition is defined as a solution that

- has good situation recognition performance,
- has acceptable response time,
- allows for manually constructed situation patterns to be adapted, and
- is applicable in real-world applications.

In this work, a new situation recognition framework by integration of CBR and approximate reasoning is proposed. Here, the actual state of ego system and its related environment is received as input. By finding relevant contexts, uncovering meaningful correlations between the characteristics, and relating the occurring situations through the events, the actual situation is identified and labeled with a descriptive type. To answer the problem statement, three problem questions should be answered.

1.2.1 Problem questions

The success of situation recognition in cognitive systems depends on knowledge representation which is still a challenge in the development of cognitive systems. To fulfill the requirement of viability, an effective knowledge representation approach should be applied to support event-discrete knowledge. This leads to the first research question, which reads as follows

> **Research question 1:** *How to model and represent event-discrete knowledge about situations for situation recognition based on approximate CBR?*

Even though knowledge representation is important to support situation recognition, a learning process is also required to adapt the knowledge. The adaptation should be according to the events resulting from new behaviors of individual human operators. Accordingly, the second research question reads as follows

> **Research question 2:** *How to improve reasoning process in approximate CBR to learn exclusive behaviors of human operators and maintain the knowledge base to support individualized situation recognition?*

The performance of the situation recognition using the proposed approach should be evaluated for a real-time application. Accordingly, the third research question reads as follows

> **Research question 3:** *How can the individualization improve the situation recognition performance?*

1.2.2 Problem objectives

To answer the research questions, three research objectives have been identified. They are noted as key milestones of this research.

The first research objective goes to address the first milestone of this research by definition and conceptualization of event-discrete situations. Knowledge base structure and its related databases, as well as data flow in reasoning process, lead to propose an effective knowledge representation approach. Therefore, the first objective reads as follows

> **Research objective 1:** *To propose and develop an appropriate knowledge representation to support situation recognition based on approximate CBR.*

Situation recognition approach should be equipped with a learning process to update the knowledge base by learning from the experienced situations. Here, the knowledge would be adapted with new experiences or behaviors of individual human operators. Thus, the second objective reads as follows

> **Research objective 2:** *To propose and develop an effective learning process to support knowledge representation approach and individualized situation recognition based on approximate CBR.*

An experimental application should be implemented to measure the performance and efficiency of the developed situation recognition approach. It is important to consider different evaluation metrics to measure the viability of the proposed approach. Accordingly, the third objective is as follow

> **Research objective 3:** *To evaluate the performance of the proposed situation recognition based on approximate CBR in terms of false alarm rate, detection rate, recognition accuracy, and recognition elapsed time.*

1.2.3 Research methodology

The research methodology undertaken in this thesis is started with a literature review on real-time situation recognition methods. Then, previous researches about personalization of situation definition and recognition for situation awareness are considered.

This work has been structured according to the three research objectives. These objectives can arrange significant milestones for investigating the problem and successfully answering the research questions. However, the nature of the objectives varies. Accordingly different methods are suitable for each objective. A brief synopsis of the applied methodology related to each objective is presented in as follows

Research objective 1 is concerned with theoretical conceptualization and definition of situation recognition problem. With this, an event-discrete knowledge modeling approach based on SOM approach is proposed and applied to CBR for the first time for formalization of the events as a sequence of scene-related changes (actions) in the dynamic changing environment. The fuzzy logic is also applied as a part of knowledge representation to handle uncertain knowledge and formalize approximate reasoning. Accordingly, the knowledge base structure will be discussed and designed to support knowledge representation of approximate CBR.

Research objective 2 is concerned with theoretical concepts of reasoning process in fuzzy SOM-based CBR. Here, a new framework for situation recognition and learning is proposed. Here, fuzzy rules are utilized for adapting and inferring the knowledge to obtain new knowledge about the actual situation. Additionally, existing methods for automatic generation of membership functions as well as feature selection for data reduction are reviewed and compared. Accordingly, membership functions generation and feature selection approaches are proposed and developed for characterizing the situations for individual human operators.

Research objective 3 is concerned with empirical results of fuzzy SOM-based CBR for individualized situation recognition. Here, a situation recognition framework based on fuzzy SOM-based CBR is designed and implemented using jCOLIBRI software [DAGCRGSRG07]. Using the open source software jCOLIBRI, the CBR process including large case bases and relational databases could be developed. In this thesis, a scenario of driving application is utilized to conduct the experimental evaluations of the proposed approach using the designed framework. However, the scenario is intended to carrying out both theoretical and experimental investigations. The proposed framework is implemented and evaluated for that application.

1.3 Thesis overview

This thesis carries out theoretical and experimental investigations with respect to the individual situation recognition using approximate CBR. This issue is addressed through the utilization of a knowledge representation approach based on SOM and fuzzy logic for modeling and structuring the event-discrete knowledge. In addition, a learning process is proposed and developed for improvement of CBR to learn new knowledge for further situation recognition. Investigations are carried out when using complex and dynamically generated situations as well as when using real-word data for real-time recognizing the situations.

A viable solution to the proposed situation recognition problem is a solution that: (1) can achieve good performance in real-time recognition of the situations, (2) can model event-discrete situations and personalize the characteristics of the situations, (3) allows for learning from new situations and adapting the knowledge base based on the new knowledge, (4) is relevant for real-world applications.

The results of this research point towards the proposed fuzzy SOM-based CBR as a viable solution to event-discrete situation recognition problem.

1.3.1 Research scopes

As argued, the results in this thesis point out that approximate CBR may be used as a viable approach for carrying out individualized situation recognition efficiently. There may also exist other techniques that can be applicable for a development of the algorithms in the formulated problem. Since the goal of this research is investigating the applicability of approximate CBR, the approach will not be compared with other comparable techniques. Naturally, in future work, it is interesting to compare the proposed approach with such techniques.

In the proposed approximate CBR, different procedures have to be progressed for knowledge retrieval, adaption, or maintenance. Accordingly, different methods could be applied for each procedure. However, selection of the methods applied to the procedures in this work is based on a literature survey on the performance of the methods in different applications and some evaluations for the experiments applied in this work. However, there is not a focused comparison among various integrations of the applied and existing methods.

In this work, the performance of the proposed approach in situation recognition is measured using different evaluation metrics: false alarm rate, detection rate, recognition accuracy, recognition elapsed time, and knowledge base workload during the learning process. In addition, 10-fold cross-validation is employed to measure the generalizability of the approach.

In this thesis, the proposed approach is applied as a framework with the ability of lane-change situation recognition and is implemented for the driving data acquired using a simulator. Here, the knowledge base could determine the applicability of the proposed approach for driving data related to 9 test drivers. However, the reasoning framework is generalized for real-time applications.

A driving situation in the target application is defined for individual human operators by a specification of definition based on the state of ego-vehicle and its related environment. However, other characteristics affecting behaviors of human drivers in different situations such as the characteristics showing humanś emotional state could be considered in the definition of situations.

1.4 Thesis outline

The rest of this thesis is structured in several chapters that outlined as follows

Chapter 2 gives the basic concepts of situation recognition and reviews existing methods for identification of the situations. In addition, there is a literature review on knowledge representation and reasoning process in fuzzy CBR.

Chapter 3 proposes a knowledge representation approach to handle event-discrete situations in fuzzy CBR.

Chapter 4 offers a reasoning process and its related procedures for situation recognition and knowledge learning for individual human operators.

Chapter 5 introduces a platform for developing the situation recognition framework. In addition, it addresses the research questions and presents experimental results related to knowledge representation and learning.

Chapter 6 summarizes and concludes the approaches, discusses research contributions and the work that has been carried out. In addition, it outlines future works.

2 State-of-the-art

This chapter consists of two sections. In the first section, the concept of situation recognition by impressing on supervision of human operators in dynamic environments is reviewed. In addition, the existing methods are introduced and discussed. The second section gives a review of fuzzy-based CBR concept as an approximate CBR and its related approaches for situation recognition. Additionally, the potential research directions of fuzzy-based CBR are addressed to advance its real-time applications.

The contents, figures, and tables presented in Section 2.2 have been prepared and submitted as the journal paper "Integration of Case-Based Reasoning and Fuzzy Approaches for Real-time Applications in Dynamic Environments: Current Status and Future Directions" [SHSed].

2.1 Situation recognition concept and methods

This chapter reviews the theoretical concept of situation awareness. Key components and related knowledge base in situation assessment are introduced. In addition, the role of situation recognition in situation awareness will be argued and modeled. Furthermore, the situation recognition methods proposed in recent years by previous researchers are reviewed and discussed.

2.1.1 Situation awareness

According to [End95], situation awareness is "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". It is an important process goes to a good decision-making and action taking in dynamic and complex environments. Situation awareness improves the human attention about the environment by relating the knowledge about the real-world or human goals and objectives. For example in a driving application, situation awareness could be reached by assessing and predicting the status of ego-vehicle, driver, and its related elements in relevant driver's goals in a driving environment.

Situation awareness is reachable in three levels [End95]: (1) perception of the elements in the environment, (2) comprehension of the current situation, and (3) projection of future status. In the first level, the status, attributes, and dynamics of relevant elements in the environment are perceived. Understanding the significance of the actual situation is the main duty of the second level. This understanding is possible by finding relations between attributes and elements in the actual situation as well as their connections with previous situations. In the third level of awareness,

a projection of possible situations by considering future actions of the elements in the environment will be given. Projection of future states could be advanced to a planning process by prediction of upcoming situations.

The three levels of situation recognition and the related components for decision-making could be modeled as shown in Fig. 2.1. Working memory is limited and includes the temporary basis of the information needed for actual situation assessment or decision-making. The long-term memory includes the stored information about experiences, goals, or criteria for decision-making.

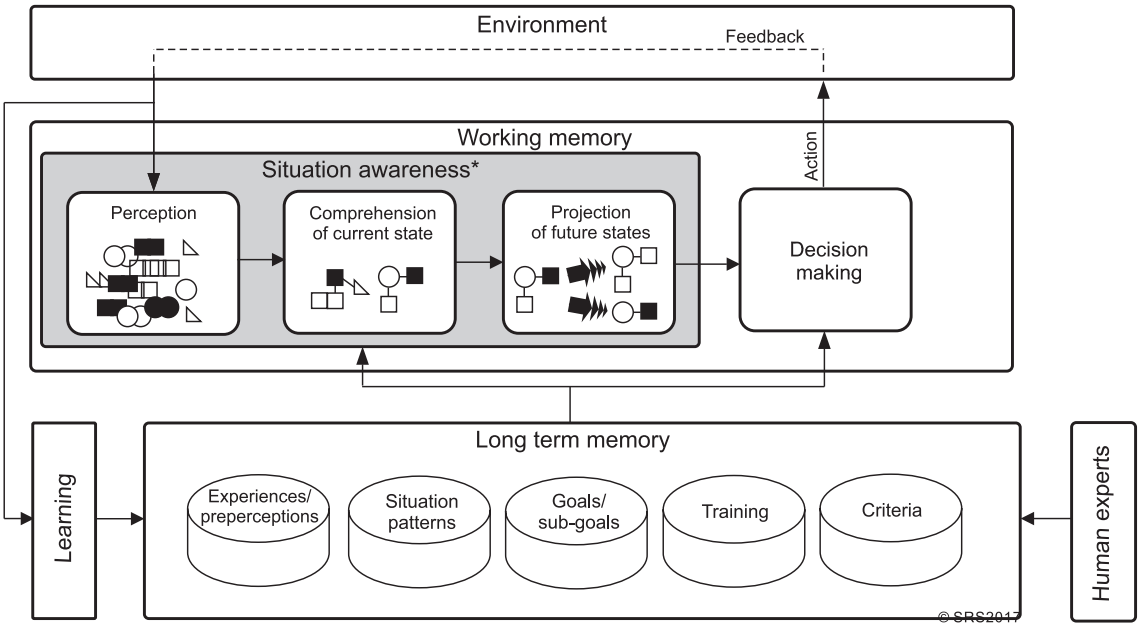

Figure 2.1: Theoretical model of situation awareness and its related components in decision-making process [SHSon]

2.1.2 Situation recognition

Situation recognition process could be addressed in the second level of situation awareness. This process can identify the significance of the current situations according to the situation patterns defined by human experts or learned through the time. The role of situation recognition in situation awareness for supervision of human operators could be illustrated in Fig. 2.2. It is supported by comprehension of current state as addressed in the second level of situation awareness. This process can identify the significance of the current situations according to the situation patterns defined by human experts or learned from new experiences.

Here, the human operator is considered as a component for semi-autonomous systems. The human operator is in interaction with the environment and cognitive

system. The cognitive system interacts with the human operator through a human-machine interface to present the information to the human about the reasoning results and to implement control operations. The perceived information is considered as the input of the process. The identified situation patterns as the output of the process go towards the projection and decision-making processes.

The projection and decision-making are processed through a high-level cognitive reasoning. A human-machine interface could be applied to support the interaction between the human operator and cognitive system. This interaction could be designed to show three types of output: (1) information about recognized situation patterns which is the output of situation recognition, (2) a decided action which is the output of decision-making, or (3) alternative actions which are the output of projection process.

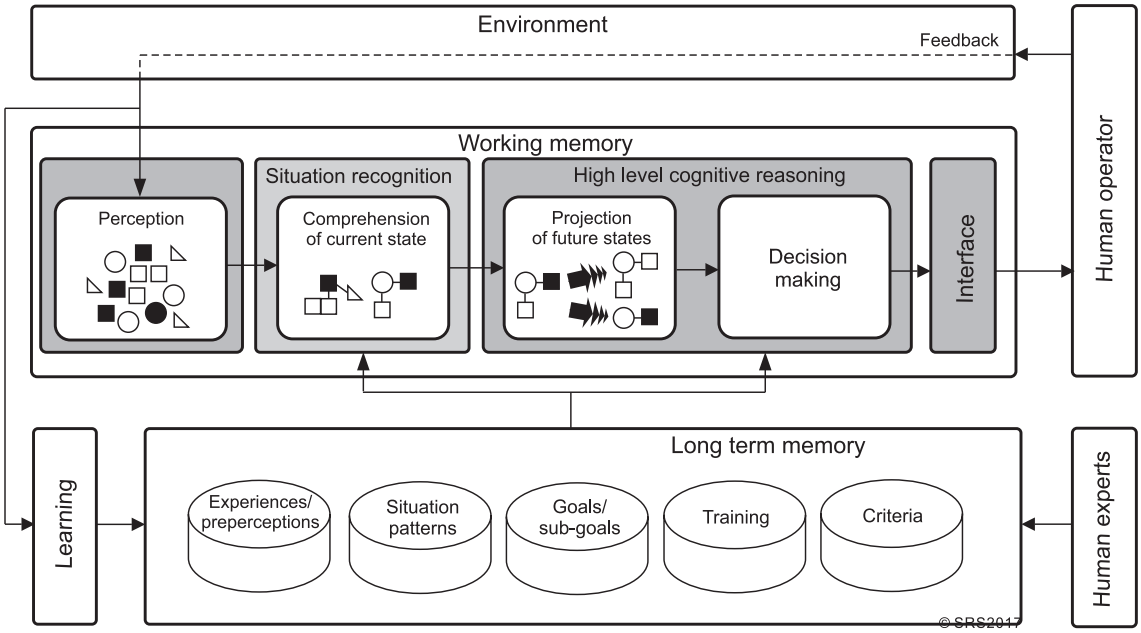

Figure 2.2: The role of situation recognition in decision-making process and supervision of human operators in dynamic environment [SHSon]

2.1.3 Situation recognition approaches

Recognition of a situation is one of the primary processes in reasoning [YDM12]. The problem of situation recognition has been addressed through different methods such as deterministic, probabilistic, and approximate reasoning. The interest in those reasoning approaches back to the very early days of artificial intelligence when problem-solving requires a high degree of intelligence . Two main components of

the reasoning approach (considered here as a knowledge-based approach) are the knowledge base and reasoning engine as shown in Fig. 2.3. The knowledge base consists of the rules or a network of existing knowledge. The inference engine works based on a logical deduction, the rules of probability theory or other theories of reasoning. By considering the knowledge base as a model, the term model-based reasoning could also be used [YDM12]. The identification results deviate from the rules of reasoning theory based on the observed states of a system.

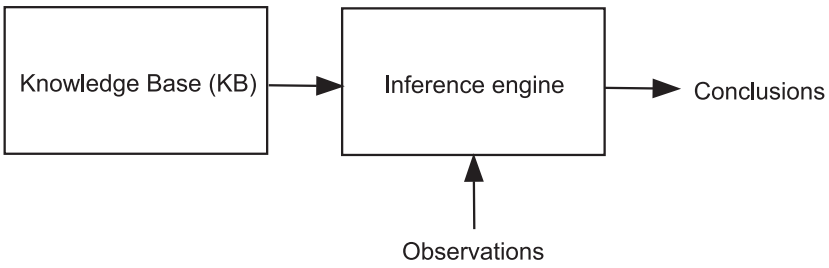

Figure 2.3: A reasoning system [Dar09]

In the following sub-sections, the main idea of these common reasoning methods will be introduced.

Deterministic reasoning

The deterministic reasoning is based on production rules to represent a situation specifications using predicate logic. A production rule is in the form, "If x obtains, derive y". The logical formula is usually generated using purely expert knowledge [YDM12]. However, the logic should be defined and precisely articulated [Ste09]. The formal logic approaches in which the knowledge about situations can be discretized and applied for deterministic reasoning [Lok10]. These approaches use rule bases for verification of the integrity and consistency of situation specifications. First-order-logic approach is an approach used for the realization of deterministic reasoning in precise system models [BBH+10].

Ontology or description Logic reasoning which can represent knowledge about a field of discourse could be considered as a deterministic reasoning. It is a formal specification of all the relevant concepts (or classes) and their meaningful associations in a given knowledge domain. Ontology could be applied to handle very large knowledge graphs for modeling and reasoning. The key concepts in ontology reasoning are entity, relation, role, and resource. The objects and instances in the domain such as "person", "man", and "doctor" are entities. The relationships between the objects such as "hasChild" are specified as relations. Roles involved in specific relationships

such as “doctor”. The resources address the characteristics associated with domain instances, such as ”name” which could be considered as properties of an object. Here, the knowledge base consists of a pair of $<T, A>$, where T is a set of background knowledge contains concepts definitions, defined roles, and axioms such as constraints, and A is a set of situative knowledge including instances of the concepts and the roles between them. Knowledge representation is based on different class/-concept constructors such as intersectionOf ($C_1 \sqcap ... \sqcap C_n$), unionOf ($C_1 \sqcup ... \sqcup C_n$), complementOf ($\neg C$), oneOf ($\{x_1\} \sqcup ... \sqcup \{x_n\}$), allValuesFrom ($\forall P.C$), someValues-From ($\exists P.C$), maxCardinality ($\leq nP$), and minCardinality($\geq nP$), where C, P, and x show the a class, role, and individual name alternatively (for more information refer to [Baa03]). As an example, a part of knowledge could be defined as follows

$$Person \sqcap \forall hasChild.Doctor.(Doctor \sqcup \exists hasChild.Doctor)\,. \tag{2.1}$$

In addition, the inference engine in ontology reasoning includes consistency, instance entailment, instance classification, instance retrieval and conjunctive queries to retrieve the individuals for which the query conditions are satisfied (for more information refer to [Hue14]).

Moreover, classical Petri net is a common approach applied for deterministic reasoning. This approach can specify and model the relationships between the situations. An advantage of considering the situations relationships is confining the search space for recognition of potential situations [BBH$^+$10]. Although Petri net provides good recognition performance, it requires a complete and exhaustive situation model including the potential situations and their relationships which are not always definable [BBH$^+$10].

Probabilistic reasoning

In this kind of reasoning, probabilistic logic is applied to express the relationships between situations using conditional probability [Yag09]. In probabilistic reasoning, the related events' probabilities could be defined using the rules.

Bayesian network is a probabilistic directed acyclic graphical model used widely for probabilistic reasoning. The knowledge base in this probabilistic reasoning is a network of knowledge with graph structure, and the reasoning engine is based on Bayes' theorem. Bayesian network uses a graph and encodes the conditional independence relationships between a set of variables $X = x_1, x_2, ..., x_n$, where n shows the number of variables. It provides the joints probability distribution $P(x_1, x_2, ..., x_n)$ over the variables. An example of the knowledge modeled by Bayesian network is as follows

$$\begin{aligned} &F_1 : P(A \mid B) \;=\; 0.6\,, \\ &F_2 : P(D) \;=\; 1\,; \end{aligned} \tag{2.2}$$

where A, B, D stand for the situations, $P(D)$ shows prior probability of the situation D before the current situation is observed, and $\mathrm{P}(\mathrm{B} \mid \overline{\mathrm{A}})$ is posterior probability of situation A after the observed situation B.

Bayesian inference computes the posterior probabilities according to Bayes' theorem as follows

$$P(B \mid A) \;=\; \frac{P(A \mid B) \,.\, P(B)}{P(A)} \,. \tag{2.3}$$

Using Bayes' theorem, the degree of belief or confidence of an occurring situation is measured based on prior observed situations.

Approximate reasoning

The approximate reasoning could be realized by applying a rule-like representation of knowledge to describe the relationships between two variables or situations. An important feature of approximate reasoning is the fuzziness and non-uniqueness of consequences of fuzzy premises.

The fuzzy logic (FL) is applied widely for approximate reasoning. It is used for reasoning on uncertain context information by representing the confidence values using membership degrees instead of probability. Each quantitative variable applied for situation identification could be defined using a fuzzy set. Each fuzzy subset is defined using a membership function. Accordingly, the knowledge base is composed of a set of admissible approximate consequences and a database of the modeling parameters of membership functions.

By considering two variables U and V and their related fuzzy sets X and Y, two approximate consequences F_1 and F_2 introduced in [Yag09] are given as follows

$$\begin{aligned} &F_1 : if\ U\ is\ A\ then\ V\ is\ B\ end\ , \\ &F_2 : U\ is\ D\ ; \end{aligned} \tag{2.4}$$

where U and V take their values in X and Y respectively, and $(A, D) \subseteq X$ and $B \subseteq Y$.

The inference engine of FL works based on the fusion operations such as intersection, union, complement, and modification.

Here, an example of two admissible approximate consequences F_1 and F_2 is given. The approximations showing the variables compatibilities are specified using the fuzzy subsets (linguistic terms) "most", "very", and "likely". The compatibility functions of those fuzzy subsets are shown in Fig. 2.3.

$$\begin{aligned} &F_1 : (most)\ students\ write\ paper\ , \\ &F_2 : Jackson\ is\ a\ student\ ; \end{aligned} \tag{2.5}$$

$$F_3 : It\ is\ (likely)\ or\ (very\ likely)\ that\ Jackson\ writes\ paper\ , \tag{2.6}$$

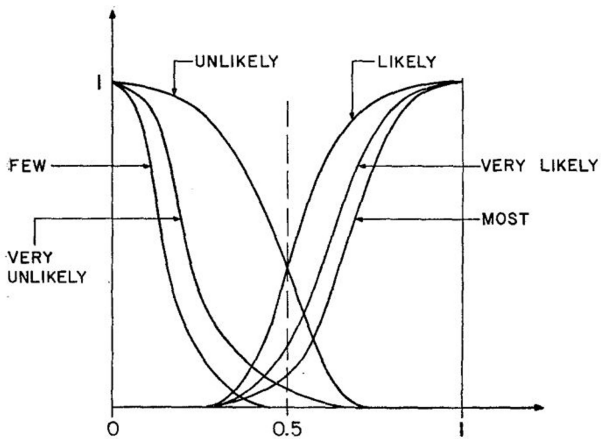

Figure 2.4: Compatibility functions (not to scale) of most, likely, very likely, unlikely, few and very unlikely [Zad75]

According to [BBH+10], FL is an appropriate approach for describing subjective contents and resolving potential conflicts between the contexts. Accordingly, to deal with vague information in many classical description logics and ontology languages, FL could be applied successfully [BS11].

2.1.4 Selection of approximate reasoning

In previous sections, different reasoning approaches for situation recognition are introduced. Those approaches have some advantages and disadvantages.

In the applications in which cognition is uncertain, or many parameters related to the environment could not be discretized, deterministic reasoning is not applicable anymore. To deal with uncertainties, the probabilistic or approximate reasoning could be applied or integrated with the deterministic reasoning for situation recognition [YDM12].

A comparison result obtained using Dempster-Shafer framework shows an underlying unity between the inferred values obtained from approximate reasoning (by applying FL approach) and probabilistic reasoning (by applying Bayesian network) [Yag09].

The probabilistic reasoning approach, the Bayesian network, is a learning-based approach relying on the models of prior knowledge. Handling the missing data

entries, avoiding over-fitting of data are two advantages of this network. However, Bayesian learning needs high computational time, and it is hard to interpret the knowledge. In addition, the probabilistic approaches may lack credibility due to the unavailability of estimations [YDM12]. It is difficult to represent the knowledge in probabilistic distributions and provide wide coverage [Uus07].

The approximate reasoning approach, FL, is a specification-based approach in which the knowledge is commonly generated based on human's expert knowledge. However, FL is responsible for knowledge representation and sharing. In this thesis, FL would be integrated with learning-based approaches for situation recognition to advance the ability of knowledge learning and deriving new knowledge.

2.2 Approximate Case-Based Reasoning

This survey reviews recent researches conducted on the application of fuzzy approaches to CBR dealing with real-time applications. Fuzzy approaches have been effectively applied for knowledge representation, feature selection, and learning in CBR. Dealing with imprecise and uncertain knowledge, generalization, mining, and learning also in combination with low computational complexity are the main advantages of fuzzy approaches used in the CBR context. This section reports new findings on the integration of fuzzy approaches with CBR. The survey result highlights the advantages of fuzzy approaches in CBR for real-time applications. It shows the current state of fuzzy-based CBR approach. In addition, it addresses the fuzzy approaches which are more operative for each operation in CBR. Those operations most contributing to the advantages of the fuzzy approach will be pointed out and detailed. Finally, some considerations of latest developments in fuzzy approaches which may be introduced as potential research directions for real-time applications are concluded.

2.2.1 Basic concepts

In recent years, some instance-based techniques in solving cognitive problems such as decision-making, situation recognition, classification have been proposed. The CBR as an instance-base approach has been successfully applied to a wide range of application-based problems such as classification, diagnosis, planning, situation awareness, and decision-making [Per14].

Case-based reasoning approach is such a technique dealing with previous experiences (cases) of the solved problem in understanding new problems. It allows avoidance of previous wrong solutions and can apply successful solutions in solving the problems. The reasoning process in CBR approach is similar to human reasoning as humans

take their past experiences into account and use them to make future decisions [GS08].

A case (defined as an episodic experience [BAF+11]) in CBR contains the knowledge about a problem description, a related solution, and the effect (a situation occurs by affecting the solution on the problem). A problem may express partly the internal structure of a system and the related environment, defined as "situation" [Söf08]. Consequently, the executed action by the system and the upcoming situation are known as "solution" and "effect" respectively [SS15]. The knowledge base applied to CBR includes the knowledge about the features, similarity assessment, and adaption conditions or rules. Case base is the main component of knowledge base utilized by CBR to store the experienced cases appending over time during the reasoning. Handling the incremental knowledge is one of the challenges in CBR which affects the system performance in real-time applications [DCGH12]. Using CBR with a large case base may cause high memory usage, high computational complexity with respect to retrieval time and memory space. These properties may reduce the system performance specially in the real-time environments with a lot of possible experiences including continues variables that need fast response time. Application of an effective case representation, an improved organization of the case base, as well as using robust methods in feature selection, case retrieval, adaption, and learning through CBR process may help to overcome the storage and computational challenges with a lot of experiences in the case of real-time realization.

Dealing with uncertainty and imprecision during knowledge reasoning is always part of a system requirement in complicated real-world applications [PY12, Kol14]. One of the current significant discussions for this issue is the integration of fuzzy approaches to improve CBR performance. According to Khanum et al. [KMJS09], the integration of fuzzy approaches and CBR has several advantages as follows

- improvement of knowledge representation,
- minimization of system computational complexity,
- isolation of using precise mathematical modeling,
- mimic of human capabilities in decision-making, and
- inferring and handling the imprecise and uncertain knowledge.

However, fuzzy approaches have some drawbacks in knowledge elicitation and learning mechanism which could be solved by integration with CBR. A system which is working only based on fuzzy approaches uses the knowledge trained in the human expert domain since it does not contain learning mechanism. However, CBR avoids this problem by providing a learning process to learn new experiences and using them for solving next problems. Fuzzy approaches could be successfully integrated

to CBR for different real-time applications with the target of planning, diagnosis, optimization, decision-making, and pattern recognition [SSA14].

Different studies in the field of CBR process have only focused on CBR applications or application of specific method used in the process. Although various fuzzy approaches have been applied successfully to CBR, no discussion on the effectiveness of fuzzy approaches in different processes of CBR for real-time reasoning could be found. The main idea of this section is to address the application of fuzzy approaches to an improvement of CBR performance for dynamic environment applications. Accordingly, this review concentrates on the effects of fuzzy-based CBR over the last recent years. By focusing on recent studies, the new findings of fuzzy approaches to enhancement of CBR will be discussed.

This section is organized as follows: In Section 2.2.2, the CBR process is discussed. In addition, the applications of fuzzy approaches to each operation of CBR are addressed in Section 2.2.3. A survey of the integration of fuzzy approaches with CBR, advantages and disadvantages of the proposed approaches, and new improvements on the related fuzzy approaches is denoted in Section 2.2.4 and Section 2.2.5. Finally, the summary and remarks are given in Section 2.3.

2.2.2 Case-based Reasoning process

Case-based Reasoning (CBR) is a methodology for solving problems in a system such as diagnosis, planning, and decision. Here considering previous experiences for solving similar problems are examined. The cycle of classical CBR shown in Fig. 2.5 is realized in four phases: retrieve, reuse, revise, and retain. Retrieve is the process of similarity assessment of the actual problem and experienced cases. In the reuse process, the solutions of retrieved cases are adapted to the actual problem to suggest a solution for the actual problem. In the next phase, revise, the suggested solution is verified in a real-world context. Accordingly, in retain phase, the actual problem and the confirmed solution can be learned and retained as a new case in the case base. Besides these four phases, two significant concepts calling knowledge representation and indexing are explained in the following sub-sections.

Knowledge representation, indexing, retrieve, reuse, revise, and retain are the main concepts defining a CBR approach. Those concepts are discussed in this section by focusing on the role of fuzzy approaches in each phase and their performance. Additionally, advantages of the fuzzy approaches in CBR operations are addressed.

Knowledge representation

Knowledge representation is defined as a process of modeling and expressing the knowledge in the computer to make them understandable and performable for rea-

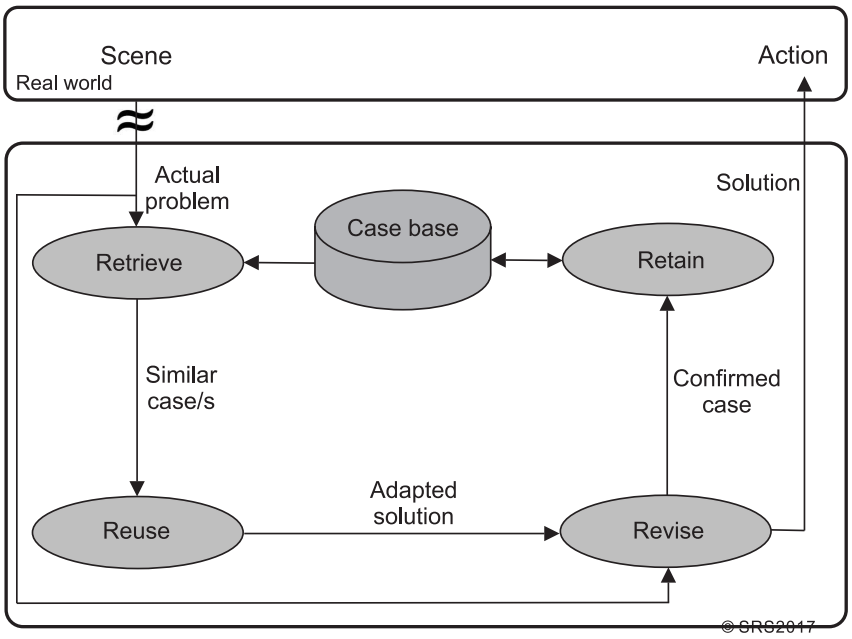

Figure 2.5: Case-Based Reasoning (CBR) process [SHSed]

soning. Knowledge in the CBR is defined as a set of cases and the related information. A case in CBR, introduced by [BKP05] as *"a contextualized piece of knowledge"* is showing an experience including a problem, the solution, and its effect.

A difference between knowledge representation approaches depends on the data structure of features and case representation forms. In this section, properties of a proper case representation and its requirements, as well as existing representation forms will be explained. The role of fuzzy approaches in the proposed representation forms and in covering the requirements will be stated.

According to Bichindaritz [Bic08] and Behbahani et al. [Mon11], utilization, adaptation, and generalization are the main properties of case representation to be examined and applied to a CBR process. Perner [Per14] demonstrates that *"generalization and abstraction make the case base more robust against noise and applicable to a wide range of problems"*. Thus, knowledge representation could ideally make a generalized case as a single experience to be reusable for a large set of problems. Accordingly, case representation should improve the dynamic learning through proper organization and classification.

For representation of a case in CBR, different approaches such as first-order logic, production rules, fuzzy production rules, frames, object-oriented, semantic networks, scripts, and Petri nets have been proposed [FxYjB+16] in recent years. Those approaches could be categorized into three major groups: feature-vector, structured, and textual representation [BKP05]. Rules and object-oriented representations as

structured approaches are two popular and contrasting schemes for representation of different knowledge [TBP12, DSS93]. Rules can represent a logical relation between characteristics and behaviors by associating actions with conditions. Moreover, feature-vector case representation could simplify case indexing for retrieval. For planning applications, a case could be defined as a planning case including the problem description, its related plan, and information about the planning decision [BKP05]. A problem recalls cases as experienced plans [MLM+05].

Case indexing

Case indexing refers to the assignment of appropriate indexes to the cases at storing time into the case base. Speeding up future case retrieval is useful. Case indexing is a learning process task. It is needed to make the case retrieval process easier and faster [BRS15,Ser10,Spa01]. According to [PY12], the indexes should be predictive, useful, and abstract enough to be retrievable at an appropriate time. Searching the cases with a significant number of quantitative features may decrease the performance of similarity assessment. Cases could be indexed by combinations of main features reflecting the important aspects of them and their applicability in various problems [Kol14]. Case indexing is also possible by considering the importance of each feature (feature weights) for similarity measurement to determine the sort of features in the generation of indices. Assigning indices could be a manual process that relies on human experts. It could also be an automatic process that relies on a feature selection process which is more applicable in unknown and dynamic environments [SS15].

Feature selection is the process of selecting a subset from given features. This process could improve data reduction process as well as retrieve and learning abilities in the systems dealing with large amounts of knowledge [DCGH12, SLS11]. A case in the case base could be retrieved by comparing the indexed features of the cases [Kol14]. As described, a case can be indexed and characterized by a set of relevant features. Redundant features may increase the case base memory size and make the similarity assessment process more complicated. On the other hand, feature selection improves the ability of case generalization by reducing the number of irrelevant features. An optimal subset of the most relevant features is found for the cases to satisfy a given criteria for similarity assessment.

Feature weighting and feature selection in CBR have been covered extensively in many types of research in recent years [SS14]. Many of the selection methods containing greedy algorithms suffer from the nesting effect. Nesting effect happens during a selection process when a removed feature could not be reselected.

Similarity assessment and retrieve

According to [Kol14], similar experiences of a situation show a strong potential with respect to situation prediction. Case similarity assessment on a case base and then recalling similar ones to solve a new problem are the main purposes of the retrieve process [Kol14]. Retrieve usually depends on a two steps search: local and global similarity assessments which are measured for each feature and total weights of case features respectively. Accordingly, the most similar cases with the best matching features will be retrieved [Kol14]. Case representation and indexing determine the complexity of retrieve process with respect to retrieval time and accuracy. In a robust retrieve process, cases should be compared at more abstract levels of representation. The similarity is evaluated in terms of indexed features and indicated using a similarity value. According to Rodriguez [RMH12], the similarity value should be normalized between 0 and 1 to be based on one scale in which large values indicate more similarity between two cases.

Adaption and reuse

A retrieved solution from the case base has to be adapted and fit to the actual problem using a reuse process. In this process, some parts of the retrieved solution need to be changed. The changes should be specified to make old solutions reusable for a new problem. According to [MLM+05], the existing adaptation methods are divided into three classes: substitution adaption, transformation adaption, and generative adaption. The substitution adaption updates only some parts of the solved solution based on the difference between the previous problems and actual problem. The transformation adaption changes the structure of the solutions. The generative adaption which is more flexible and applicable in various applications, suggests a new solution by inferring from similar experienced solutions. The fuzzy approaches applied to CBR for reusing the solutions are based on generative adaption.

Learning, revise, and retain

Learning process in classical CBR comprises two processes: revise and retain. Through the revise process, correction of an adapted solution is verified. Knowledge maintenance which could be triggered periodically, conditionally, or offline during a pause in reasoning is the primary task of retain process. However, the learning process may carry out a set of tasks such as data reduction to support the retain process.

The knowledge in CBR is upgraded by learning new experiences on the new solved problems. In the retain process, new experiences should be evaluated for the learning process. Different learning methods for short term and long term learning have been proposed by the researchers. The learning process in a system uses the knowledge

of the actual problem, the revised solution, and new status of the system obtained from the real-world (the effects of the solution on the problem). The process of adding new experiments in the case base is denoted as incremental learning [CG11, Hül07]. Incremental learning may increase the size of the case base during the learning [BAF+11]. The CBR approach generally requires large memory space and computational time in retrieve and inference processes. Accordingly, maintenance improves the performance of CBR by applying different approaches for storing the new cases, updating the case base, data reduction, and removing the irrelevant features of the cases [BAF+11, Kol14].

In [Smy98], maintenance is defined as a process with the abilities of eliminate the out-of-date, redundant, or inconsistent cases, as well as grouping or generalizing the cases to reduce redundancy. According to [Smy98], *"effective maintenance relies on an ability to model a complete performance (competence and efficiency)"*. The purpose of competence-based maintenance is saving the cases with low reachability rate and high coverage rate. Reachability rate is defined as a set of cases which can be replaced by the target case. Coverage rate is defined as a set of problems which could be solved by the target case and shows the generalization capability of a case. In contrast, efficiency-based maintenance covers special efficiency of the system such as reaching to the high retrieval time by decreasing the case base size or removing competence-critical cases. Thus, a maintenance process by considering both aspects of competence and efficiency could enhance the CBR performance.

2.2.3 Applied fuzzy approaches to CBR

In this section, an overview of the application and efficiency of the fuzzy approaches in integration with CBR is given. The results of a literature review on the integrated fuzzy approaches with CBR for real-time application domains are presented in Table 2.1. According to the review results, fuzzy set (FS), FL, fuzzy rough set (FRS) theory, fuzzy cognitive map (FCM), and fuzzy decision tree (FDT) have been applied successfully to the CBR process. These approaches and their applications in CBR are addressed in the following sub-sections.

Fuzzy sets

Fuzzy sets are used for the representation of variables with a degree of possibility and necessity without sharp boundaries. Improving the tolerance for imprecision by transforming the quantitative values to qualitative values is an advantage of FS. A quantitative value of a continues variable could be encoded using fuzzy set membership functions.

A survey on the application of FS in CBR (Table 2.1) shows that FS could be successfully applied for representation of knowledge. On the one hand, the value of

Table 2.1: Application of fuzzy approaches in CBR [SHSed]

Approach	Knowledge representation	Indexing	Retrieve	Reuse	Revise and Retain
FS	[BDPB13, CG11, LYH12] [KMJS09, FDE08, RG02] [DFE08, KKC05, LH09] [CFD10, CLBV08, Nik08] [AEGT04, DPDRJ11, GAG06] [Xio11, XMZ13, DHP02] [AL13, SKK05, KPAA14] [RADMV09, ZHQ$^+$15, ZRK11] [HZL$^+$14]	[Kol14]	[AEGT04, CG11, LH09]	[RG02]	[CLBV08, GAG06, CSW01] [AIA10, Kol14]
FL	[AEGT04, ZRK11]		[BAF$^+$09, BAF$^+$11, RADMV09] [XMZ13, KMJS09, DHP02] [Xio11, LH09, AL13] [ZRK11, LYH12, CNSZ11]	[LYH12, DHP02, PY12] [FRDC07, Kol14, XMZ13] [LH09, DPDRJ11, Nik08] [YBG12, RG02, KPAA14]	[SKK05, XMZ13, KMJS09] [Xio13]
FRS		[AHY11, SLS11, FDE08] [RJB$^+$14, Wan06, ZLC14]			[ZHQ$^+$15, AHY11, CSW01] [FRDC07]
FCM	[KML14, DCGH12, Pap11] [DPDRJ11, SS14, GWL13]	[DCGH12, MJ12, MSA14] [VR08, SLS11]		[CLBV08, GWL13]	[CLBV08, GWL13]
FDT	[LYH12, Pap11]	[CFD10]			[SHGA14, SE11]

a single continues variable could be encoded for several sets with different grades. This ability could improve the tolerance for imprecision. On the other hand, different values of a single variable may belong to one set with different grades. This ability enhances the case generalization when a case could belong to different problems. Application of FS in different domains is addressed in the following sub-sections and summarized in Table 2.2.

Knowledge representation using FS In [FDE08,DFE08,KKC05], two different methods for structuring the data in structural representation of cases in CBR are applied: crisp and fuzzy. The results show that using FS to represent data values, the case base size without any significant difference in retrieval accuracy can be reduced compared to the crisp values. The fuzzy set could be useful in a generalization of the cases according to their features values. In addition, fewer cases in the case base will be stored, since fuzziness encodes and adapts the features values more effectively [DHP02]. However, handling incomplete and uncertain knowledge as well as structuring the continues variables are reported as advantages of FS by [FDE08], [DFE08], and [KKC05].

The FS approach has been successfully applied for case representation to CBR in real-time financial activity prediction [LH09] and coordinated action selection in robot soccer [RADMV09]. In [FDE08, DFE08], a fuzzy histogram is utilized to represent visual perception of the agents. The obtained results by those contributions show that the position of the objects in scene could be effectively approximated using fuzzy parameters rather than conventional crisp set. In addition, Khanum et al. [KMJS09] and Biswas et al. [BDPB13] and [HZL$^+$14] applied FS for representation of the cases to simplify the measurement of facial action and rainfall prediction respectively. Representation of imprecise and uncertain values of features will be feasible by application of fuzzy approaches [LYH12]. In the proposed CBR approaches, fuzzy sets are defined based in the human expert domain for representation of data.

In [RG02, CG11], FS is applied to CBR for data structuring in object-oriented representation and feature-vector representation of cases respectively. The object-oriented representation of the cases with complex structure can represent each case at multiple levels of abstraction [AS14]. This helps to enhance the indexing ability for various similarity measurement (see Section 2.2.2).

Indexing using FS The researchers showed that applying FS for representation of feature values could be used as a complement for fast indexing the cases [Kol14]. This approach can be used to represent adaptable values of features and therefore transforms the quantitative features into a few discrete groups. Therefore, FS could decrease the complexity of retrieval time and memory space.

In addition, application of FS improves the ability of multiple indexing [JL95]. A solution in a single case could contain a potential solution for various problems as its features may be encoded with different membership grades.

Retrieve using FS As discussed, knowledge representation and indexing affect the performance of the retrieve process. In [CG11,LH09,AEGT04], FS is applied for case representation. Intersection areas between the defined membership functions are considered for similarity assessment in real-time applications.

Reuse using FS In [RG02], a case presented using object-oriented model is able to describe the knowledge using possibility distributions. In this presentation, FS is also applied for modeling the adaption constraints.

Learning using FS As reported by [CLBV08, GAG06, CSW01], case base maintenance could also be performed successfully by applying FS for representation of features value and generalization of the values in feature level. Qualitative values could facilitate the generalization of similar situations by assigning the same FS to the features with similar value [RADMV09]. This kind of generalization in feature level overcomes the problem of knowledge discrimination since the similarity of the generalized features for two cases is considered for generalization. In addition, by presenting the cases as fuzzy rules (see Section 2.2.2), a generalization in case level could be realized.

Applying FS as feature structure, automatic generation of membership functions would be a part of case base maintenance. In recent years, different algorithms have been applied for refining the membership functions by space partitioning. Fuzzy c-means and probabilistic c-means are suitable methods utilized for different application domains [CATU10]. The design modeling parameters for membership functions are determined using fuzzy c-means algorithm [PS04].

Fuzzy logic

Fuzzy logic is an approximate reasoning approach for managing and inferring the vague information by utilizing a fuzzy set of values and suitable set of fuzzy inference rules. Fuzzy inference system relies on the inference rule to match the value and obtain adapted results [PS13]. Therefore, a rule base containing a set of fuzzy rules plays an important role in FL.

Fuzzy logic is applied for knowledge representation and complex decision processes to give the ability of knowledge inference by handling linguistic terms. According to Table 2.1, FL is integrated with CBR mostly for case retrieval. In addition, it is applicable for representation of knowledge, reuse and retain processes.

In the application of FL, the fuzzy rules could be generated initially by human expert knowledge or automatically according to the learned cases in the case base. Fuzzy logic benefits the advantages of FS since fuzzy rules apply FS for structuring the values. Furthermore, FL overcomes classical logic in improving the CBR performance by inheriting human reasoning process in handling the situations with uncertain knowledge. However, extraction of fuzzy rules from human expert domain or automatically from case base is an important issue in the application of FL. Use of FL in different application domains are addressed in the following sub-sections and summarized in table 2.2.

Knowledge representation using FL Fuzzy if-then rules could be applied as a structural representation of cases in CBR. By using this representation, CBR benefits the advantages of adding, modifying, and deleting the cases [AEGT04, XMZ13]. The fuzzy rules could be generated by expert knowledge or using the experienced cases. According to [XMZ13], representation of the cases as fuzzy rules could improve the computational efficiency with respect to retrieval time and memory space, since fuzzy rules could be applied as case matching mechanism for retrieving the similar cases.

Retrieve using FL In recent years, FL has been effectively applied for similarity assessment [BAF+09, BAF+11, LYH12]. In this application, a set of fuzzy if-then rules is organized for the CBR to measure the similarity between two cases. Using fuzzy rules and fuzzy induction, the concepts of similarity and its possibility by considering uncertain features could be established [DHP02]. Case retrieval using fuzzy matching techniques may be relevant [LYH12]. In addition, by considering additional information about the risk of features and feature weighting, the most appropriate cases could be retrieved using fuzzy inference. This process called "suitability process" in [CNSZ11], uses the information about risks (the danger involved when an experienced value is used for a new problem) and weights of features to adapt the fuzzy inference for retrieval. According to [CNSZ11], CBR has a good features weighting by considering risk concepts in fuzzy similarity rules. Additionally, the most suitable cases could be retrieved compared to the conventional CBR in integration with ANFIS and K-NN approaches. Using the proposed method in [CNSZ11], the most suitable cases which are sometimes different from the most similar ones would be found.

In [AIA10, XMZ13] an integration of FL and CBR has been applied for case retrieval. In the applied approach, cases are defined as rules based on fuzzy set. The membership functions and fuzzy rules for fuzzy inference are generated by human experts and during the design process. The proposed approach in [AIA10] can estimate the hazard with acceptable accuracy. Moreover, Khanum et al. [KMJS09] applied FL to CBR for an inference based on the indexed features and retrieving the most similar

cases for pattern recognition. In addition, [BAF+09, BAF+11] applied FL to CBR to propose decision support system in health sciences.

In [Xio11, AL13, XMZ13], FL is proposed for assessing the similarity between the cases. The evaluation results using different case studies stress the efficiently of FL in retrieve process. As discussed in Section 2.2.2, fuzzy rules could be applied for case representation or case matching mechanism. On the other hand, fuzzy rules could be justified in the human expert domain with the purpose of similarity assessment. Fuzzy rules could be generated from the learned case using evolutionary techniques such as genetic algorithms [Xio11].

Reuse using FL Fuzzy logic has been widely applied to CBR for case adaption as shown in Table 2.1. In this application, fuzzy rules are generated from the experienced cases in the case base for fuzzy inference. According to [PY12], case retrieval by means of fuzzy matching techniques and case adaptation by using the concept of production rules could be relevant. According to [LYH12], case adaptation by using the concept of gradual rules would be relevant. In [DHP02], the fuzzy rules for automatic adaption, selection, and a combination of relevant features are defined in the human expert domain. Nevertheless, evolutionary techniques have been applied successfully for generalization and optimization of fuzzy adaption rules [YBG12].

Learning using FL Research results on the suitability of fuzzy approaches for case base maintenance [SKK05, Xio13] indicate that FL could be effective by generalizing and explaining numerical/quantitative values to linguistic terms, and supporting competence-based and efficiency-based strategies.

Fuzzy rough set

The theory of Fuzzy rough set (FRS) is used as a well-known approach for identification and recognition of common patterns in data, especially in the case of uncertain and incomplete data [AEGT04, MSA14]. Using FRS theory, a fuzzy approximation of a set including the lower and upper approximations of the original set is definable [CLBV08, GAG06, MSA14, Per14]. Accordingly, a reduced set of features giving the same information as the original set could be generated.

In recent years, an increasing interest in an integration of FRS theory with CBR for feature selection process can be observed [AHY11]. Through feature selection process, the features importance and the most relevant features of the cases could be measured. As it is shown in Table 2.1, the most two applications of FRS are case indexing and learning. The discrimination between these two applications is not easy because features importance considered for indexing would be determined

through a learning process. Accordingly, the application of FRS in both indexing and learning processes is addressed here.

In [Wan06, FDE08], FRS is applied for feature weighting and reduction in CBR. The researchers proved that FRS in integration with CBR could successfully measure feature weights for similarity assessment and removes redundant features. If the features contain quantitative values, FRS does not need discretization [RJB+14]. Consequently, it may effectively be feasible and avoid losing information as well as the nesting effect. In addition, an integrated FRS with CBR is proposed in [SLS11] to measure the relevance of each feature and evaluate the number of features sufficient for the given data. The data should be discretized first because FRS can only handle nominal values. In [SLS11], a floating search method and threshold strategy as criteria for feature selection are used. The proposed approach in [SLS11] has been presented for maintenance and learning processes. The researchers show that the fuzzy approximation operators could be represented as membership functions to improve the robustness of FRS with respect to misclassification of the patterns and enhance the features value reduction [ZTCW10]. In [Wan06], FRS uses the information of experienced cases as well as the cases clusters generated by Self-organizing Map. The evaluation results on the performance of the proposed approach by [Wan06], conventional CBR, and ANN for short-term load forecasting show that the proposed approach could improve the prediction accuracy. In [ZLC14], FRS theory has been integrated with CBR for feature weighting in a real-time estimation of product particle size.

Fuzzy cognitive map

Fuzzy cognitive map (FCM) is an inference engine that uses the expert knowledge or the knowledge achieved from the real-world in the form of fuzzy rules [CLBV08]. Features in cognitive maps could be represented using fuzzy values. Fuzzy rules in FCM are derived from knowledge using knowledge extraction algorithms such as fuzzy decision tree, association rule-based methods, and neuro-fuzzy methods [DCGH12, MSA14]. The fuzzy rules in FCM emphasize the relation between features (key factors of the system). The structure represents the significance of the system. The degree of relationship between two features is usually fuzzified. In addition, FCM supports dynamic involving feedback, where any changes in the features impress other features [CLBV08]. Approximate reasoning and dealing with incomplete knowledge have been addressed as the advantages of FCM in integration with CBR.

Knowledge representation using FCM In [DPDRJ11], it is shown that FCM could be applied as a complete representation of structured knowledge in CBR for modeling the knowledge and supporting the planning process in complex systems.

This fuzzy approach is a relational model for representing the causal knowledge between concepts and inferring the concepts using fuzzy inference. Here, an experienced problem in CBR is realized as a concept in FCM. A relation between the concepts is represented with a link showing the effects of experienced solutions. By considering the existing relations which could be organized dynamically during a learning process or initially by human experts, FCM can infer and evaluate the effects of the possible solutions [Kol14]. Most of the studies applying FCM such as [KML14, DCGH12, Pap11, DPDRJ11] benefit its relational structure between the concepts as case representation.

According to [DPDRJ11, Nik08], FS could successfully represent the values in FCM. In the proposed FCM, a case could be easily applied for decision-making process and planning in different application domains such as engineering, control, robotic, medicine, and business [CLBV08].

Reuse using FCM In [GWL13, CLBV08], FCM has been applied to CBR for recognition of goal situations in unknown environments. In this approach, a goal is verified and described using the connection between related sub-goals. In this application, sub-goals extracted from a goal rule are verified using FCM inference and the rules which are connecting the sub-goals [GWL13]. This verification is done using fuzzy probability generator and fuzzy inference. The ability of FCM in real-time inference for dynamic environments improves the performance of goal-driven decision-making. The experimental result on robot navigation goal detection shows that FCM can be applied effectively for real-time applications [CLBV08].

Learning using FCM Moreover, FCM has been applied by Ganganath et al. [GWL13] and Chandana et al. [CLBV08] in mobile robot navigation. In the proposed FCM, FS is considered for structuring the features in a cognitive map. Fuzzy rules in FCM emphasize the relation between features. These rules could be defined by human experts or derived successfully using knowledge extraction algorithms such as fuzzy decision tree and neuro-fuzzy [CLBV08]. It is shown that integration of FCM with CBR allows a high detection accuracy in the application of goal detection.

Fuzzy decision tree

The fuzzy decision tree is a popular rule-based reduction approach which is applied to different applications for learning and classification. This approach has been effectively applied for reducing the case base size using a group of adaption rules generated by FDT [DPDRJ11, ZTCW10, DHP02]. In [CG11], an approach for learning according to the distance matrix weight and genetic algorithms for optimization of the fuzzy decision tree and selecting optimal fuzzy terms is introduced.

Knowledge representation using FDT By integration of FCM with CBR, a knowledge extraction method could be applied to cognitive map to extract the experienced cases in the form of fuzzy rules [LYH12]. In [Pap11], three rule-based extraction methods using fuzzy decision tree, fuzzy association rule, and neuro-fuzzy based algorithm have been proposed and evaluated. The evaluation results address that fuzzy decision tree gives the rules with a good level of interpretability for decision-making.

Learning using FDT According to [SHGA14, SE11], FDT could also be integrated with CBR for knowledge base maintenance with the rules defined by human experts. The maintenance is progressed by learning, case reduction, feature weighting, clustering the cases based on the evaluated feature weights, and adapting the rules using FDT. Finally, the representative cases are selected in each cluster with low reachability and high coverage rate.

2.2.4 Actual status of fuzzy-based CBR approaches

The typical problems solved by integration of fuzzy approaches and CBR could be related to six application groups: classification, prediction, reasoning, decision-making, planning, and fault diagnosis (see Table 2.2).

According to the information given in Table 2.2, the main focus of the applications is for classification, prediction, and decision-making. In most of the proposed approaches, FS is applied for case representation. Fuzzy similarity rules and FL are considered for similarity assessment. The similarity rules could be defined by human experts or generated from the case base by other methods such as genetic algorithm (GA), ant colony optimization (ACO) algorithm, and particle swarm optimization (PSO) algorithm. The adaption could also be realized by applying fuzzy adaption rules defined by human experts or neural network. In the learning process, different clustering methods such as fuzzy c-mean, fuzzy k-nearest neighbor (k-NN) are applied to the CBR. In addition, FRS and FL are widely applied to CBR for the case and feature reductions.

Current status of the proposed fuzzy-based CBR cycle could be defined in five types as listed in Table 2.3. Fuzzy-based CBR-1 and -2 are generated by giving a focus on knowledge representation changes. The integrated approaches, fuzzy-based CBR-3 and -4, are obtained by adjusting the retrieve process. The presented fuzzy-based CBR-5 is afforded by modifying the learning process. In the following sub-sections, the defined approaches are graphically described according to their main focuses. The main advantages and disadvantages of the proposed approaches are given in Table 2.3.

Table 2.2: Application of fuzzy approaches in CBR for different problems solving [SHSed]

Application	Reference	Knowledge representation	Similarity assessment	Adaption	Learning and indexing
Classification	[FDE08]	FS	K-means clustering	-	FRS
	[DFE08]	FS	-	-	FL
	[KKC05]	FS	ACO	-	-
	[GWL13]	FCM	-	-	FCM
	[CFD10]	FS	FDT, GA	-	-
	[SLS11]	-	-	-	FRS
	[ZTCW10]	-	-	-	FRS
	[CNSZ11]	-	Fuzzy similarity rules	-	Considering weights and risk
	[KMJS09]	-	-	-	FL
	[Xio11]	FS	FL, GA	-	-
	[XMZ13]	FS	FL, GA	-	-
	[AEGT04]	FS	FL	-	-
	[PY12]	-	-	FL	-
	[Xio13]	-	-	FL	FL for case reduction
	[CSW01]	-	-	-	FRS
	[CATU10]	-	-	-	Fuzzy c-means
	[AHY11]	-	-	-	FRS for case selection and weighting
	[ZHQ+15]	-	Fuzzy similarity rules	-	FRS
Prediction	[LH09]	FS	FL	-	FS
	[RADMV09]	FS	-	-	-
	[KMJS09]	FS	-	-	-
	[BDPB13]	FS	-	-	-
	[HZL+14]	FS	-	-	-
	[ZLC14]	-	FRS	-	No feature reduction
	[DHP02]	FS	-	FL, human experts	-
	[LH09]	FS	FL, GA	-	Fuzzy KNN
	[FRDC07]	-	-	FL, radial basis function neural network	FL, FRS
	[XMZ13]	FS	FL, human experts	-	-
Reasoning	[RG02]	FS	-	FL, constraints modelled by FS	-
	[CG11]	FS	FS	-	-
	[SKK05]	FS	-	-	FL
Decision-making	[LYH12]	FR, FDT	FCM, FL	FL	-
	[Pap11]	FDT, FL, neuro-fuzzy	-	-	-
	[CLBV08]	FS, FDT, neuro-fuzzy	-	-	FS for generalization
	[GAG06]	FS	FCM	-	FS
	[BAF+09]	-	FL	-	-
	[BAF+11]	-	FL	-	-
	[AL13]	FS	FL, human expert	-	-
	[YBG12]	-	-	FL, PSO	-
Planning	[KPAA14]	FS	-	-	
	[DPDRJ11]	FS, FCM	-	FL	-
	[Nik08]	FS, FCM	-	FL	-
Fault diagnose	[Kol14]	-	-	FL	FS

Fuzzy-based CBR-1

The approaches FL and FCM have been popularly integrated with CBR to represent the cases in the form of structured and planning cases respectively, where the cases are defined as causal rules. On the other hand, FS has been admitted effectively for structuring the features in a case and using fuzzy-valued feature instead of crisp-valued features for quantitative features by most of the studies.

Applied changes to CBR cycle and the related knowledge base to support knowledge representation approaches can be displayed graphically via Fig. 2.6 and Fig. 2.7.

According to the approaches presented by [FDE08, DHP02, DFE08, KKC05] for knowledge representation, a fuzzy-based CBR approach could be illustrated in Fig. 2.6. As it is shown in Fig. 2.6, to apply FS for structuring the features, a fuzzification process is applied to generate fuzzy data by discretization of the continues values. In addition to the original case base, a fuzzified case base including the converted fuzzy data is applied to the knowledge base. Moreover, a database of fuzzy sets to save modeling parameters of fuzzy membership functions is used for fuzzification process. According to the reasoning cycle illustrated in Fig. 2.6, cases in the fuzzified case base are used by retrieve process and are generated through retain process.

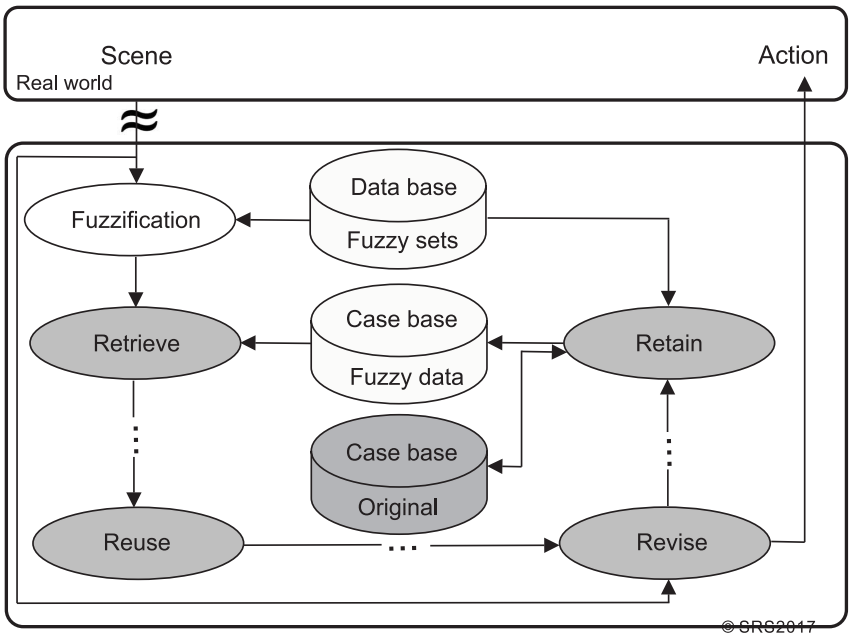

Figure 2.6: Fuzzy-based CBR-1: CBR and the knowledge base when the fuzzy sets are applied as data structure for features [SHSed]

Fuzzy-based CBR-2

In Fig. 2.7, a CBR cycle including its related knowledge base when fuzzy rules are applied for knowledge representation is proposed. This approach inherits the techniques justified in [AIA10, ZRK11, XMZ13] for application of fuzzy rules to infer the knowledge for reuse process.

Here, the knowledge base consists of the original case base, a fuzzy rules case base presenting the cases as rules, and a database of modeling parameters of fuzzy membership functions for fuzzification and inference. The proposed approach could integrate retrieve and reuse processes by applying FL engine to infer the cases and obtain a final solution for the actual problem.

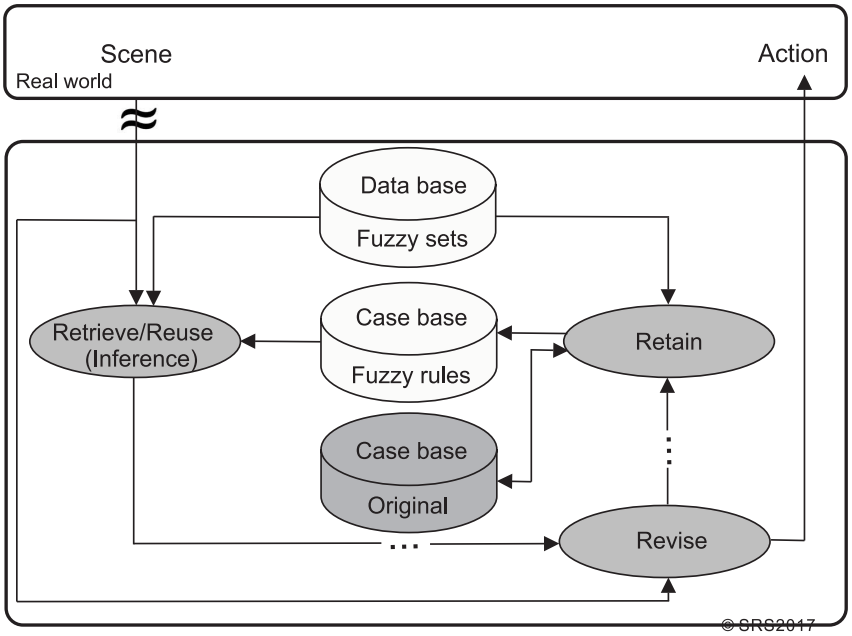

Figure 2.7: Fuzzy-based CBR-2: CBR and the knowledge base when the cases are stored in the form of fuzzy rules as knowledge representation and a solution is reused by inferring the whole cases [SHSed]

A survey on [PY12, YBG12, DHP02, LH09] shows that FL in integration with CBR is effective for reuse process. Reuse could be done by inferring the cases through a fuzzy engine. Here, cases should be defined as fuzzy rules to prepare the knowledge for inference. The fuzzy rules defined for adapting the retrieved solutions in the proposed CBR have been generated in the human expert domain up to now.

Fuzzy-based CBR-3

As discussed in previous sections, the performance of similarity assessment process strongly depends on knowledge representation and learning. According to Table 2.2, FL has been used effectively for similarity assessment by application of fuzzy similarity rules and assessment.

According to the research results presented by [BAF+09, BAF+11, LYH12] for similarity assessment in CBR, a new presentation of fuzzy-based CBR could be illustrated in Fig. 2.8. As it is shown in Fig. 2.8, the knowledge base includes a rule base of fuzzy similarity rules and a database of fuzzy membership functions. Here, fuzzy similarity rules are applied to compare the similarity of the cases with the actual problem. The similarity assessment rules in FL could be defined in the human expert domain or automatically from the learned cases in the case base.

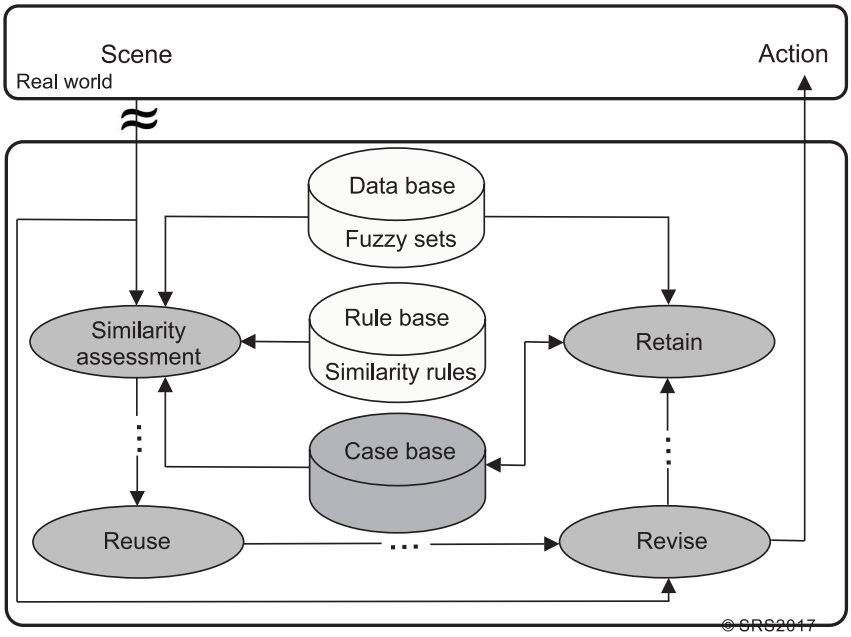

Figure 2.8: Fuzzy-based CBR-3: CBR and the knowledge base when the similarity assessment is based on similarity rules and fuzzy membership functions [SHSed]

Fuzzy-based CBR-4

Retrieve process in the defined fuzzy-based CBR-3 could be improved by application of additional information for cases similarity assessment. Here, fuzzy-based CBR-4 as shown in Fig. 2.9 is an improved version of fuzzy-based CBR-3 according to the technique introduced by [CNSZ11]. Here, similarity rules benefit a database of

feature risks and weights to find the most suitable cases instead of the most similar ones.

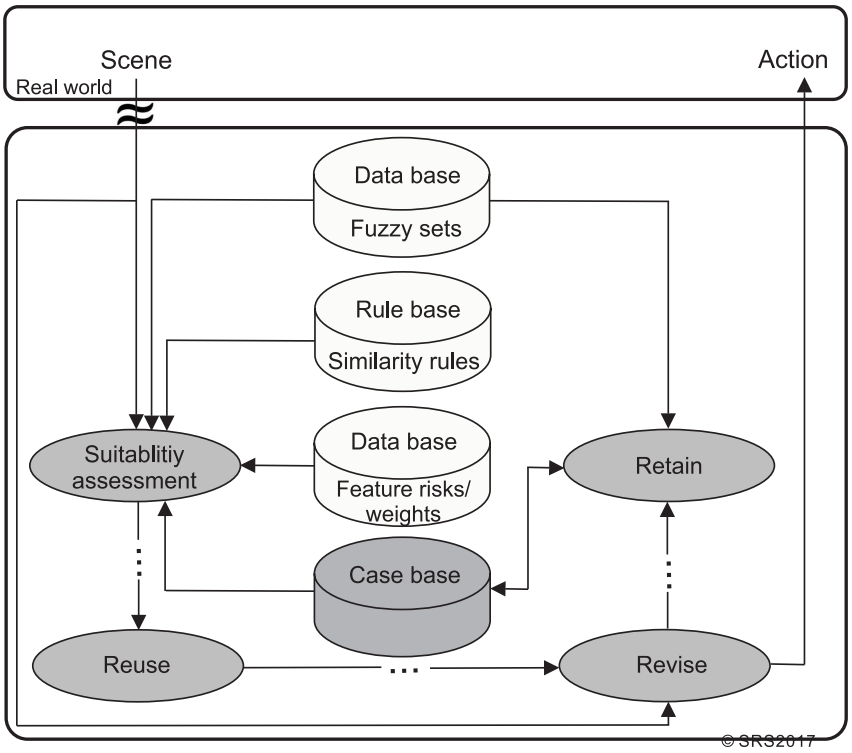

Figure 2.9: Fuzzy-based CBR-4: CBR and the knowledge base when suitability of the cases is measured instead on similarity [SHSed]

Fuzzy-based CBR-5

According to the Table 2.2, the approaches FL and FRS are mostly used for learning. Moreover, a survey on [BAF+11, Kol14, Wan06] shows that application of FS for structuring the features gives the ability of knowledge generalization for retain process. In addition, feature selection and weighting for indexing the cases and retrieve are realized mostly using FRS.

The CBR approach and its related knowledge base when FS is offered for data structure, are proposed in Fig. 2.6 as fuzzy-based CBR-1. To apply feature selection and weighting functions to the learning process, the knowledge base needs to be improved as illustrated in Fig. 2.10. The proposed fuzzy-based CBR-5 is supported with a database of features weights which are determined during learning. Accordingly, the application of case or feature reduction functions is necessary for retain process. This information could be initialized and updated through online

learning or by human experts [Smy98]. Feature weights and key features are used to improve the accuracy of similarity assessment.

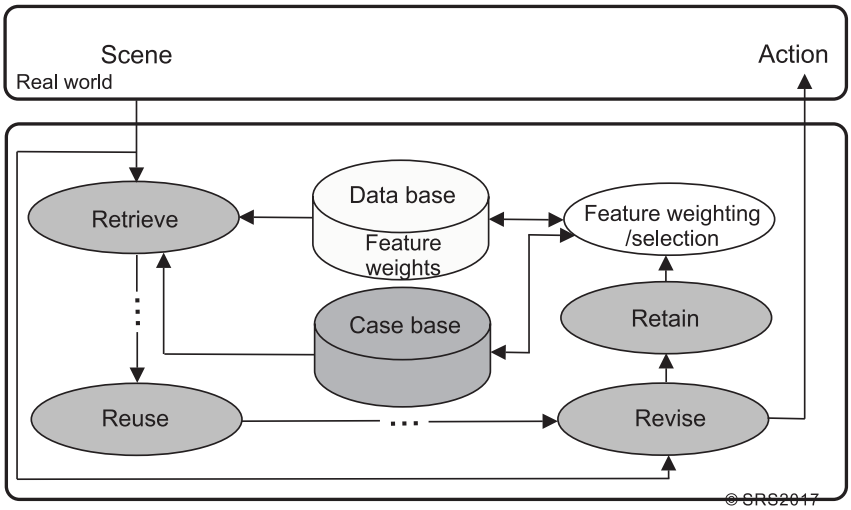

Figure 2.10: Fuzzy-based CBR-5: CBR and the knowledge base when learning process is equipped with feature selection and weighting functions [SHSed]

2.2.5 Improvement potential

By considering the advantages and disadvantages of the proposed approaches listed in Table 2.3, a new fuzzy-based CBR approach could be proposed. In addition, the new improvements in fuzzy approaches done by researchers which have not been already applied to the CBR will also be discussed and suggested in this section for further improvements of CBR process.

A proposal for fuzzy-based CBR

As shown in the previous section, different fuzzy-based CBR approaches have been proposed to improve the performance of CBR in different application domains. In this section, a new fuzzy-based CBR approach is proposed by considering the advantages and disadvantages of the approaches listed Table 2.3. The proposed approach is shown in Fig. 2.11 and is specified as follows.

To benefit the advantage of case indexing in fuzzy-based CBR-1, a fuzzification process is applied to the proposed CBR process. Hence, a database of fuzzy sets is added to the knowledge base. To support case adaption with knowledge inference as applied to fuzzy-based CBR-2, the cases will be represented using fuzzy rules in

Table 2.3: Proposed fuzzy-based CBR approaches [SHSed]

Approach	Main focus	Advantages	Disadvantages
Fuzzy-based CBR-1	Knowledge representation	• Dealing with uncertainties • Case generalization • Case indexing ability	Definition of fuzzy sets using human experts
Fuzzy-based CBR-2	• Knowledge representation • Adaption	• Dealing with uncertainties • Cases generalization • Knowledge inference • Working with causal knowledge	• No indexing • No similarity assessment • Definition of fuzzy sets using human experts
Fuzzy-based CBR-3	Retrieve	• Application of fuzzy similarity rules	• Definition of fuzzy sets using human experts • Definition of similarity rules using human experts
Fuzzy-based CBR-4	Retrieve	• Application of fuzzy similarity rules • Using key features for similarity measurement	• Definition of fuzzy sets using human experts • Definition of fuzzy similarity rules using human experts
Fuzzy-based CBR-5	Learning	Automatic feature selection and weighting	

the proposed approach. However, using fuzzy rules as case structure depends on the application domain. The retrieve and reuse processes should work separately to benefit the advantages of similarity assessment in pattern recognition. A knowledge inference could be realized for reuse process by handling the cases as fuzzy rules. Additionally, to avoid the disadvantage of dependency to human experts knowledge, the database of fuzzy sets could be upgraded through learning.

The proposed fuzzy-based CBR could also benefit the advantage of fuzzy-based CBR-4 in suitability assessment instead of similarity assessment. Here, a database of feature risks and weights is considered and initialized. However, the usage of this ability depends on the application domain.

Similar to fuzzy-based CBR-5, utilization of a feature selection and weighting in retain process gives the opportunity for data reduction and finding key features for case indexing and retrieve in the proposed fuzzy-based CBR.

Support of fuzzy-based CBR with some new extensions of fuzzy approaches

Fuzzy sets Although FS affects the performance of CBR in different processes, estimation of membership functions is still a problem. Finding appropriate membership functions is an important issue in the application of FS. The membership functions could be defined by relying on human expert knowledge. Although there

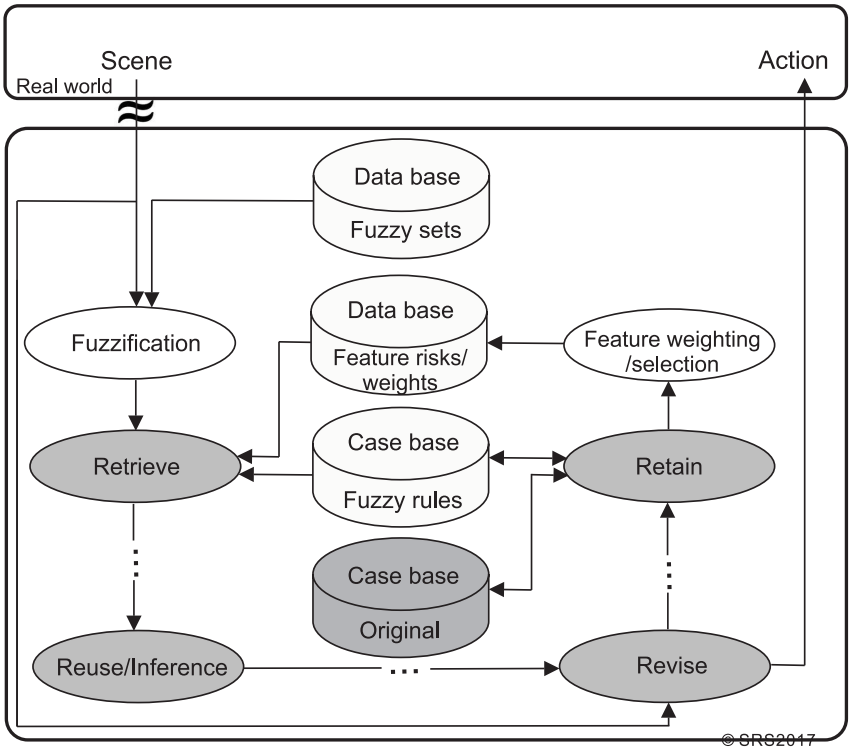

Figure 2.11: Proposed fuzzy-based CBR approach and the knowledge base when the cases are stored in the form of fuzzy rules as knowledge representation and a solution is reused by inferring the similar cases [SHSed]

are some methods in automatic generation of functions from data, they have not been applied to CBR approach yet. Dynamic adjustment of membership functions for continues variables may enhance the accuracy of functions in a dynamic environment. The adjustment depends on the changes in distribution or repetition of experienced values during the run-time. It is improved by adding or removing fuzzy sets and by tuning membership function parameters. Therefore, automatic generation and optimization of functions are important processes in the learning process.

Different generalizations and extensions of fuzzy sets introduced by [RMH12] could also be used instead of classical FS in CBR approach to handle vague and imprecise information simultaneously and model the ordinary fuzzy sets.

In the application of FS, when the system hesitates about several values for a linguistic variable, the fuzzy linguistic approach is very limited. To improve the elicitation of linguistic information, hesitant fuzzy linguistic term sets introduced by [RMH12] could be applied. It is effective mainly when the system hesitates among several values to assess fuzzy set variables. This method which is applied successfully for

decision-making purposes could be considered for representation of knowledge in CBR.

Fuzzy rough set Although FRS is an effective approach for dimensional data reduction, it is usually ineffective for large case bases. Fuzzy boundary region, fuzzy discernibility matrix, and fuzzy positive region are new FRS-based feature selection approaches presented in [JC11] which have not been previously applied to CBR. An evaluation result shows that the suggested approaches in [JC11] could find smaller subsets and speed up the searching process compared to the FRS, although the classification accuracy for all the approaches is the same. In addition, the fuzzy discernibility matrix could return the smallest set of relevant features [JC11]. The performance of CBR approach may be improved by application of the new FRS-based feature selection approaches in a real-time application.

Moreover, neighborhood-based multi-granulation rough set (NMGRS) introduced by [VDSC15, LQL12] is suggested to improve knowledge discovery from information systems with various domains of attributes. The evaluation results show that NMGRS could improve the knowledge discovery from various data sets compared to FRS.

Additionally, some implementations of MapReduce, iMapReduce, and traditional methods combined with MapReduce model in FRS (introduced by [ZWLP14]) are recommended for large-scale data mining and knowledge discovery. An evaluation result shows that knowledge acquisition using MapReduce has better performance in term of computational time compared to the traditional rough set for large-scale data mining in run-time applications.

Fuzzy cognitive map To assign linguistic terms for FCM nodes relationships within complex problems, belief-degree-distributed fuzzy cognitive maps (BDD-FCMs) introduced by [RHM11] would be useful. It is especially applicable if the human expertsj́udgment has to be done for the prediction. The evaluation result shows that the proposed technique is advantageous to model and analyze complex dynamical systems [RHM11].

2.3 Discussion and conclusion

The approach CBR has been widely proven through the literature as a useful reasoning technique for different real-time and complex dynamic systems. However, application of an effective case representation, an improved organization of case base, as well as using robust methods in feature selection, case retrieval, adaption, and learning in CBR process may help to overcome the low accuracy, storage and

computational challenges with a large amount of experiences and uncertainties in the case of real-time applications.

This chapter presented a survey on the application of fuzzy approaches to improvement of CBR performance for different problems. In recent years, fuzzy approaches have been successfully integrated with CBR to overcome these challenges in classification, prediction, decision-making, reasoning, planning, and fault diagnose in complex dynamic systems. As a critical overview, the application of fuzzy approaches is speedily growing up specially in the fields of classification, predictions, and decision-making.

The literature review on the application of fuzzy approaches at CBR shows that the fuzzy approaches are effectively applied for CBR process. It is addressed that FS, FL, FRS, and FCM are the applicable fuzzy approaches in CBR.

The main improvement of CBR is related to the case representation by applying FS and organizing a fuzzy case base. The main motivation of using FS approach is its ability in representing adaptable values of features, case generalization, low computational complexity in retrieval time and memory space.

Representation of cases with fuzzy rules improves the ability of CBR for modeling the casual knowledge. It could also facilitate case retrieval and adaption using the concept fuzzy similarity assessment and inference. A motivation of using FL approach is its high flexibility in representing the knowledge and criteria in fuzzy rules using incomplete knowledge for inference. In addition, FL gives high understandability in the definition of similarities in the human domain. In this thesis, FL is applied for knowledge representation approach to CBR (this application supports the thesis objective 1).

Moreover, by reviewing the related approaches, a new fuzzy-based CBR approach with improvement potential is proposed. The proposed approach fuzzy-based CBR may benefit the advantages of the approaches applied previously for real-time applications in situation recognition and learning abilities. In this thesis, the new proposed fuzzy-based CBR is applied for the development of a new real-time situation recognition platform (this development and its evaluation support the thesis objectives 2 and 3).

In addition, FRS is effectively applied for feature weighting and similarity assessment as well as feature/case reduction in the learning process. On the other hand, FRS approach offers a high flexibility in the presentation of features with imprecise values. Production of an optimal subset of features, representation of continuous data, and no requirement for values discretization are other advantages of FRS in case indexing process.

The literature review result shows that conventional extensions of fuzzy approaches have been successfully integrated to CBR to solve some CBR drawbacks such as

dealing with uncertainty, redundancy of data, retrieve, and learning. There are some new extensions of fuzzy approaches which are considered in different research domains but have not been applied to the CBR. These extensions of conventional fuzzy approaches introduced partly in this section may be a useful trend to overcome conventional approaches and expert dependency.

3 Knowledge representation for event-discrete situations

The knowledge base in CBR consists of a case base and a set of related databases. Case base as the main component of CBR plays an important role in situation recognition process. An effective representation of cases could strongly affect CBR performance and improve the accuracy and computational time. In this section, a new approach for structuring and representing the cases is introduced. In the proposed approach, SOM is applied for modeling the cases and FL is integrated for structuring the case in the case base. In this way, the CBR approach may deal with incremental knowledge for situation recognition. To support knowledge representation approach and related processes for case base maintenance, a set of databases are defined.

In this chapter, a new approach for knowledge representation is presented to support individualized situation recognition. The CBR approach is applied to realize situation recognition by remembering human operator experiences for supervision. Major challenges of learning in individualized situation recognition are related to modeling and representation of new experienced knowledge. To deal with these challenges and to improve the learning ability, the proposed CBR is combined with SOM and FL approaches.

This chapter is divided into different sections. Application of classical CBR for situation recognition and its procedures are discussed in Section 3.1. Section 3.2 introduces SOM approach and its abilities for modeling event-discrete knowledge. Application of SOM in representation of cases in CBR is addressed in Section 3.3. Section 3.4 demonstrates the application of FL in representation of knowledge and the requirements for fuzzy structuring the data in case base. Finally, a summary of the chapter is given in Section 3.5.

The contents, figures, and tables presented in this chapter have been prepared for publication as the journal papers "fuzzy SOM-based Case-Based Reasoning for individualized situation recognition applied to supervision of human operators" [SHS17b] and "learning and representation of event-discrete situations for individualized situation recognition using fuzzy SOM-based CBR" [SHSon], and partly published as the conference papers [SS15, SHS17a].

3.1 Situation recognition using Case-Based Reasoning

This reasoning approach applies previous experiences for solving new problems. Similar experiences could be found by applying different similarity assessment techniques. Knowledge inference techniques may be operated to reuse previous experiences for the new problem. In addition, new problems and their solutions could also be learned and applied to the knowledge base. An incremental learning is applied to CBR to store experiences in a knowledge base called case base. An experience exposing a problem and its solution is organized as a case in the case base.

As it is shown in Fig. 2.5, the main processes of CBR are retrieve, reuse, revise, and retain. Retrieve process assesses the similarity of a new problem and the experienced problem, and return the most similar case/s. Reuse process adapts the retrieved cases with the actual problem to suggest an appropriate solution. In revise process, the suggested solution is applied to real-world. A new case may be confirmed for a new experience of the problem, the solution, and its result. Accordingly, the confirmed case is learned and stored in the case base for further usages.

The term and construct situation is used to model a real-world scene [Söf01]. It expresses the internal structure of a considered system, as a part of the real-world. A specific combination of sensor data and related combined characteristics is understood as a situation representing the actual state of the system and the related environment [Söf08]. A CBR approach applied for solving a special cognitive problem introduces a situation and the performed actions as "problem" and the executed upcoming situation as "solution".

A situation is recognized into different levels. According to [RC12], the main three recognition levels are based on features of the situation, connections between the features, and the relation between situations. These three levels could be illustrated according to Fig. 3.1. In the first and second level, a complex situation is identified by retrieving the most similar experienced situations. Finally, the retrieved situations are identified in the third level by comparing the related situations in adapted cases.

Realization of these recognition levels is possible by applying a proper case representation approach to support features, connection between the features, and relation between the situations for each case. Since situation recognition using CBR is accomplished by learning from the experiences, employing a suitable knowledge representation approach to support incremental knowledge is required. In the next section, the CBR approach is improved by application of an effective case representation approach.

3.2 Situation-Operator Modeling (SOM)

Situation-Operator Modeling is an approach for modeling event-discrete structural variable systems. Using SOM, behaviors of cognitive operators in a dynamic en-

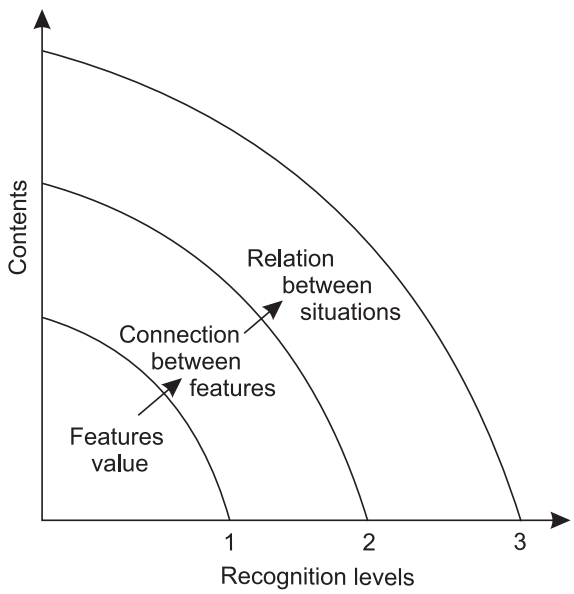

Figure 3.1: Situation recognition levels [SHS17b]

vironment could be modeled, learned, and planned for further cognitive reasoning. This approach could interpret and model the events in a real-world as a sequence of scenes, related causes (actions), and their effects [Söf08]. In the following subsections, the key elements in SOM approach as well as case structure by applying SOM approach are briefly presented.

According to the [YDM12] defining different relationships between situations, the SOM technique is focusing on temporal sequence relationship which may be realized using actions. In this type of relationship, each situation may occur before or after another situation. This relationship may be realized using the actions.

The approach SOM defined in [Söf01] for the first time, is a technique for modeling event-discrete situations specially in dynamic environments with structural changes. Using this modeling technique, an event (a change in the environment) could be represented using a sequence of scenes, related causes (actions), and their effects. These concepts are modeled using two items, situation and operator, to represent scenes and actions alternatively. Accordingly, an event is defined as a sequence of initial situation, performed operator, and upcoming situation. This sequence is illustrated graphically in Fig. 3.2. Situation S_{i} is defined as a set of characteristics C with various hybrid data types and quantitative values, as well as relations R between characteristics [Söf08]. The characteristics are quantities illustrating the features of a scene as a part of the real-world. The relations indicate the internal structure of a situation by using function-oriented connections between characteristics as defined in Eq. 1.1. The relations may be effected by changing the values of characteristics.

Additionally, an operator models an action which changes scenes. The item operator O applies function f to transfer situation S_{i} to $S_{\mathrm{i+1}}$ with possibly a new set of characteristics' values and the relations. The operators considered by SOM are modeling the active actions.

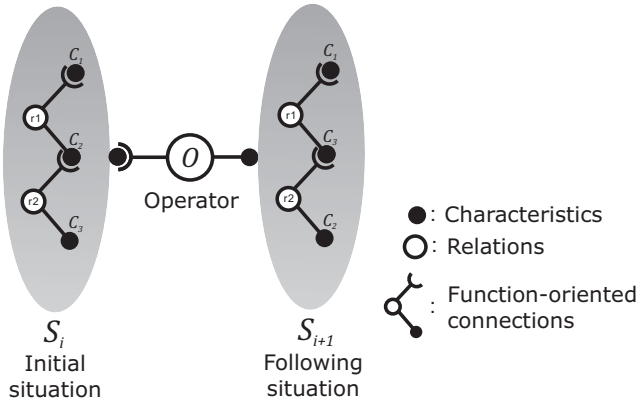

Figure 3.2: Graphical illustration of the terms situation and operator [Söf08]

By modeling the events happening continuously, a long sequence of situations and operators representing the environment changes may be generated. A sequence of situations and operators is illustrated in Fig. 3.3, where S_1 and S_{n} show initial and final situations. The operators may change the situations by changing the characteristics values or relation between the characteristics. The sequence may cause a specific situation pattern.

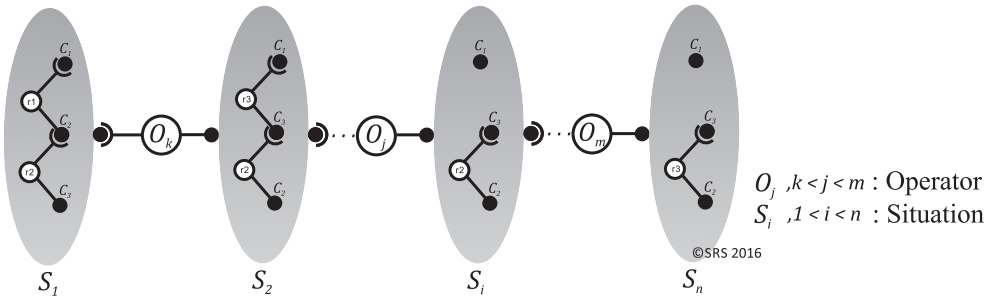

Figure 3.3: Graphical illustration of a sequence of situations and operators [SHS17b]

By application of SOM for representation of event-discrete experiences in CBR, a sequence of events could be considered as a set of cases while each case is representing an event. Accordingly, the case $Case_{\mathrm{j}}$ where j is the case index, is defined using a 3-tuple $(S_{\mathrm{i}}, O, S_{\mathrm{i+1}})$. Assuming a closed and time-invariant arrangement (given system structure), S_{i} and $S_{\mathrm{i+1}}$ are two situations of an infinite set of situations, and O is an operator.

In addition, several operators could be summarized and introduced by a meta-operator. By applying this representation model, the interaction with an agent can be explained easily. The approach SOM is able to generate an action space by using a set of situation-operator sequences resulting from propagation of alternative actions from an initial action. Generation of action spaces is one of the main properties of cognitive systems due to the related ability for planning, evaluation, etc.

The model identified by SOM approach could represent the characteristics and relations between the combined characteristics in a situation as well as giving a sequence between situations. Therefore, an SOM-based CBR could support all three recognition levels introduced in the previous section (see Fig. 3.4). Here, SOM is applied to CBR as an operational approach representing the case in technical systems. A case in SOM-based CBR gives a sequence of an initial situation S_{i}, the operator O, and the upcoming situation $S_{\mathrm{i+1}}$. Accordingly, experiences are stored in case base as a list of SOM-based cases.

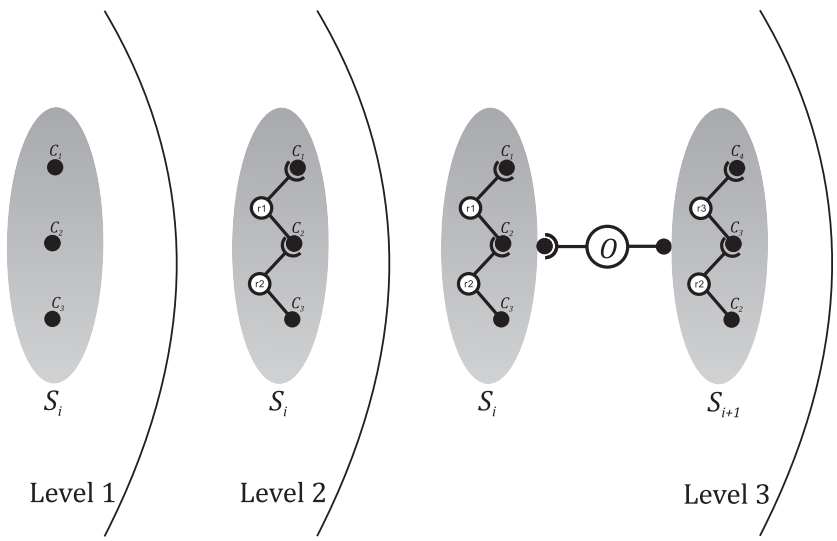

Figure 3.4: Situation recognition levels based on SOM concept [SHS17b]

3.3 SOM-based cases

The case base in the proposed approach is defined as a fuzzy object-oriented knowledge base. Cases are represented by an object instantiated from the class diagram.

A situation in SOM is defined using a set of characteristics as $S_{\mathrm{i}} = (c_1,\ c_2,\ ...,\ c_n)$, where n is the number of characteristics defining a situation. An operator is defined using a set of characteristics as $O = (o_1,\ o_2,\ ...,\ o_m)$, where m is the number of

characteristics defining an operator. A case is a combination of initial situation S_i, operator O, and the upcoming situation S_{i+1}. Situations and operator are two object types defined using a set of characteristics with various data types.

Scheme of the case base is described with a class diagram shown in Fig. 3.5 where c_j ($j = 1..n$) addresses the j-th characteristic of class situation in which n is the number of characteristics defining a situation. In addition, o_r ($r = 1..m$) shows the r-th characteristic of class operator O in which m is the number of characteristics defining an operator. Variables p_f ($f = 1..q$) shows the f-th operator of class meta-operator P in which q is the number of operators defining a Meta-operator.

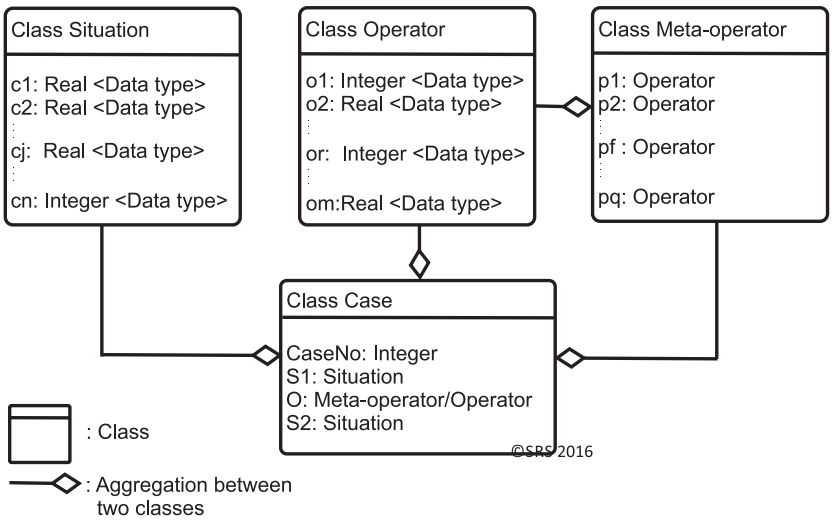

Figure 3.5: Class diagram of SOM-based CBR [SHS17b]

3.4 Fuzzy SOM-based cases

In different application domains, FL is applied to deal with uncertain knowledge and to formalize approximate reasoning. In this approach, quantitative values of characteristics expressing detailed parameters of the environment are transferred to fuzzy sets. Therefore, a similarity assessment could be developed by applying fuzzy values instead of using discrete or continues values of parameters. This makes the cases usable in various situations [SS16b]. Applying this approach could decrease the complexity of knowledge in the case base by generalization of comparable cases. Additionally, the experienced knowledge used for supervisory tasks of human operators can be expressed in linguistic terms.

Fuzzy logic could be integrated with the proposed CBR for structuring the knowledge as well as knowledge induction for retrieving and adapting the experiences for

actual situation [SS16b]. In this section, applications of FL in the proposed approach are highlighted.

3.4.1 Case fuzzification

As previously mentioned, a situation is defined using a set of characteristics which may be associated with the continuous and quantitative values changing over time. Modeling the situations using quantitative values of the characteristics contradicts generalizability of the knowledge. Accordingly, correctness of the transformation of the real scalar and precise values to the corresponding imprecise values is important for knowledge representation.

In addition, application of various linguistic modifiers to characterize the situations may cause considerable effects in interpretation and identification of the situations. Those effects are detectable in situation recognition performance for supervision of human operators as shown in Fig. 2.2. Therefore, using optimal number of linguistic modifiers improves individualization of situation recognition.

In this contribution, fuzzy sets are applied to handle imprecision and uncertainties. A fuzzy set is defined as a 2-tuple (α,μ) where α, $\alpha \in [0, 1]$, is an input crisp value and μ is a membership function. Accordingly, the crisp value x of a characteristic is converted to its corresponding fuzzy values, x_f, using the membership vector

$$x_f = [\mu_{L1}(x), \mu_{L2}(x), ..., \mu_{Ln}(x)], \tag{3.1}$$

where the vector elements represent the membership degrees to the membership functions $\mu_{Li}(x)$ of the linguistic modifiers Li ($1 \leq i \leq n$), where n is the maximum number of modi

ers.

Linguistic modifiers are defined for the characteristics with quantitative values. Definition of an optimal number of linguistic modifiers is an important task of fuzzification process. These numbers are not equal for different characteristics and should be defined or measured individually. The designed parameters of membership functions should be stored in an extra database of fuzzy design parameters (DB_{FDP}).

3.4.2 Fuzzy case base

By application of FL in SOM-based CBR, a case could be defined by rules to support event-discrete events with inference process. A sequence of situations and operators shown in Fig. 3.6 is addressed by a set of cases as follows

$$\begin{aligned} &Case_1 : if\ (S_1)\ \&\ (O_1)\ then\ S_2;\ end; \\ &Case_2 : if\ (S_2)\ \&\ (O_2)\ then\ S_3;\ end; \\ &Case_3 : if\ (S_3)\ \&\ (O_3)\ then\ S_4;\ end. \end{aligned} \tag{3.2}$$

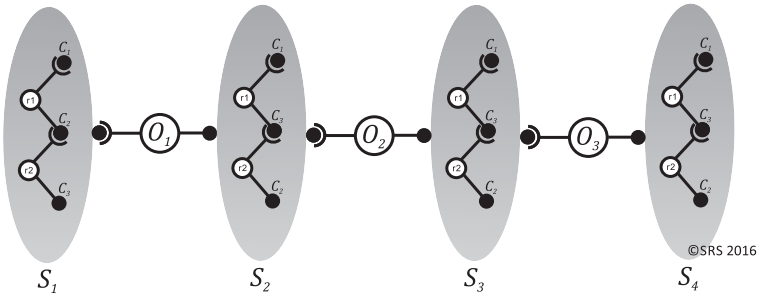

Figure 3.6: A sequence of situations and operators modeled by SOM [SHS17b]

Characteristics with crisp values are stored based on the fuzzy sets to define linguistic terms. Using linguistic terms instead of quantitative values of parameters is suitable for generalization of experiences. By applying fuzzy sets, the scheme of the case base is described with a class diagram as shown in Fig. 3.7.

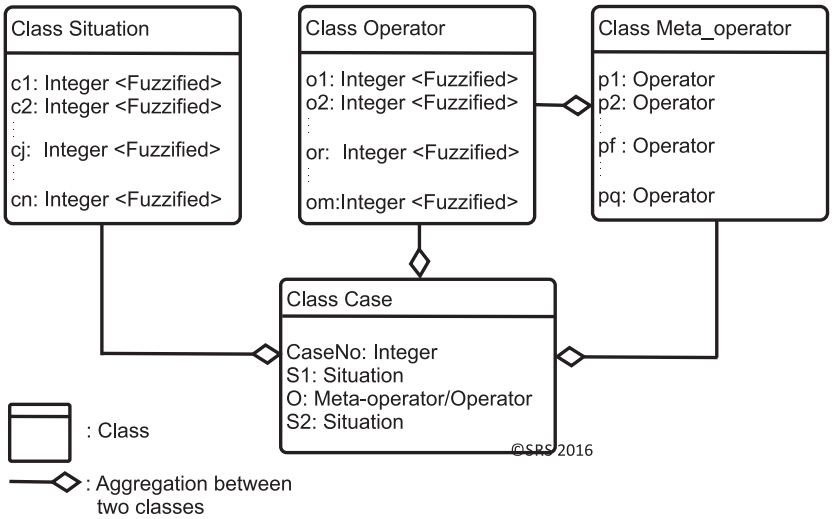

Figure 3.7: Class diagram of SOM-based case by applying fuzzy sets [SHS17b]

By fuzzification of characteristics values, a case (an event modeled by SOM) could be represented using a fuzzy rule where antecedent (if-part) indicates initial situation and operator, and consequent (then-part) of the rule demonstrates the upcoming situation.

By applying fuzzy logic, the l-th case in the case base is retrieved as a fuzzy rule defined as Eq. 3.3, where $1 \leq l \leq k$, k is the number of cases in a case base. Variables c_j $(1 \leq j \leq n)$ addresses the j-th characteristic of a situation in which n is the number of characteristics defining a situation. Accordingly, $S_i.c_j$ and $S_{i+1}.c_j$ indicate the j-th

characteristic of initial situation and the next situation respectively. In addition, o_r $(1 \leq r \leq m)$ represents the r-th characteristic of the output operator O in which m is the number of characteristics defining an operator. Variables A_j, B_j, and D_r are the related values assigned to the characteristics of initial situation S_i, upcoming situation S_{i+1}, and operator O respectively. Each characteristic of a case may be assigned to multiple membership functions. Therefore, each case may be split into multiple instances and represented through multiple fuzzy rules. To preserve the original cases from the environmental or behavioral changes by overwriting the fuzzy cases, it is suggested to apply an additional fuzzy case base ($CaseBase_{fuzzy}$). In the $CaseBase_{fuzzy}$, each fuzzy rule is retrieved as a case.

$$
\begin{array}{l}
Case_l: \\
if\ (S_i.c_1 = A_1\ \&\ ...\ \&\ S_i.c_j = A_j\ \&\ ...\ \&\ S_i.c_n\ =\ A_n) \\
\&\ (O.o_1 = D_1\ \&\ ...\ \&\ O.o_r = D_r\ \&\ ...\ \&\ O.o_m = D_m) \\
then \\
\quad S_{i+1}.c_1\ \leftarrow\ B_1; \\
\quad \vdots \\
\quad S_{i+1}.c_j\ \leftarrow\ B_j; \\
\quad \vdots \\
\quad S_{i+1}.c_n\ \leftarrow\ B_n; \\
end.
\end{array}
\tag{3.3}
$$

Through the reuse process of CBR, an adapted situation could be inferred from the cases retrieved in the form of fuzzy rules.

3.4.3 Similarity assessment

The characteristics may have different contributions in definition of a situation. These contributions present the importance (impact factor) of each characteristic in identifying the situation. Therefore, similarity of the cases could be measured based on the impact factor of each characteristic in identification of situation patterns. The impact factors of the characteristics identifying the s-th situation pattern could be shown as a vector

$$W_s = [w_{s1}, w_{s2}, ..., w_{s\imath}, ..., w_{sn}], \tag{3.4}$$

where n shows the number of characteristics. The element $\mathrm{w}_{s\imath} \in \mathbb{R}$, $0 \leq \mathrm{w}_{s\imath} \leq 1$ where $1 \leq \imath \leq n$ indicates the impact factor of the $\imath$-th characteristic.

These factors may be specified generally or may depend on human operator's priorities or preferences in defining that situation. Moreover, the applied factors could

effectively be used for indexing the cases and speed up of case retrieval based on the importance characteristics in an indexed case base ($CaseBase_{indexed}$). However, the importance factors are retained in a database of indexing features (DB_{IX}) for similarity assessment and retrieve (as described in Chapter 4).

In retrieve process of case-based reasoning, k-NN is widely and successfully applied for real-time applications [CMS14]. In this work, a derivative of k-NN classifier by consideration of weighted features and applying a modification of local similarity assessment (as detailed in the following part) is used for case similarity assessment.

As mentioned in Section 2.2.2, the similarity of two cases is measured based on local and global similarity assessment. Similarity of two comparable characteristics x and y is calculated using the similarity of individual features (local similarity) as follows

$$\begin{array}{ll} Similarity_{local}(x,y) = 1 - \frac{x-y}{\propto} & if\ 0 \leq |x-y| \leq \propto, \\ Similarity_{local}(x,y) = 0 & otherwise \end{array} \tag{3.5}$$

where $\propto$ shows the slope of the function. This function is inspired by the function introduced in [RG02] and modified by applying impact ratio for each characteristic. The application of this ratio improves the flexibility of the function by considering the effects of each characteristic on the similarity. This function returns a value between 0 and 1 showing the similarity of the parameters values. Higher values indicate more similarity. This function is applied to all characteristics defining a situation. The total (global) similarity of the situations S_α and S_β is calculated as follows

$$Similarity_{total} = \frac{\sum_{j=1}^{n} (w_j \times Similarity_{local}(S_\alpha.c_j, S_\beta.c_j))}{\sum_{j=1}^{n} w_j}, \tag{3.6}$$

where c_j ($1 \leq$ j $\leq$ n) addresses the j-th characteristic of a situation in which n is the number of characteristics defining a situation. Accordingly, $S_\alpha.c_j$ and $S_\beta.c_j$ indicate j-th characteristic of the actual situation and sub-goal situation respectively. As well, w_j shows the impact factor of the j-th characteristic.

3.5 Summary

In this chapter, a CBR approach is applied for development of situation recognition. The knowledge base and knowledge representation approach are improved by

application of fuzzy SOM approach to model event-discrete situations experienced by human operators. The proposed fuzzy SOM-based CBR is addressed as an individualized approach by applying personalized knowledge of exclusive definitions of experienced events. In the next chapter, a new framework is developed to support the proposed knowledge representation approach for individual situation recognition.

4 Improvement of reasoning process for individual human operators

To support the proposed approach for knowledge representation in approximate CBR, new processes as well as a new knowledge base are required. Situation recognition process should be provided with new operations to apply the fuzzy SOM-based cases. The knowledge base is defined as a collection of case base and other databases to store related information. It should support data exchange between the processes.

Main processes of CBR are considered for the tasks of situation recognition and learning the experienced events with the target of individualized situation recognition in fuzzy SOM-based CBR. In the following sections, these two tasks and the related procedures are more detailed.

The contents, figures, and tables presented in this chapter have been prepared for publication as the journal papers "Fuzzy SOM-based Case-Based Reasoning for individualized situation recognition applied to supervision of human operators" [SHS17b], "learning and representation of event-discrete situations for individualized situation recognition using fuzzy SOM-based CBR" [SHSon], "Automatic fuzzification of continuous variables for approximate reasoning" [SHNSon], and partly published as the conference papers [SHS17a, SS16b, SS16a].

4.1 Fuzzy SOM-based CBR

According to the literature review results presented in Section 2.2, a fuzzy-based CBR is proposed in Section 2.2.5 to benefit the advantages of previously applied fuzzy-based CBR approaches for real-time applications.

According to the suggested fuzzy-based CBR shown in Fig. 2.11 and by considering the offered knowledge representation approach based on SOM and FL, a new framework is proposed in this section. The dynamical flow of the proposed CBR-based situation recognition scheme is shown in Fig. 4.1.

This framework is presented to support two important tasks for individual supervision of human operators: situation recognition and learning. These two tasks are detailed in the following sections.

4.2 Situation recognition

Situation recognition is supported using two main processes: retrieve and reuse. However, some additional procedures are applied to support fuzzy SOM-based CBR. The procedure of situation recognition is detailed using Algorithm 1. The knowledge

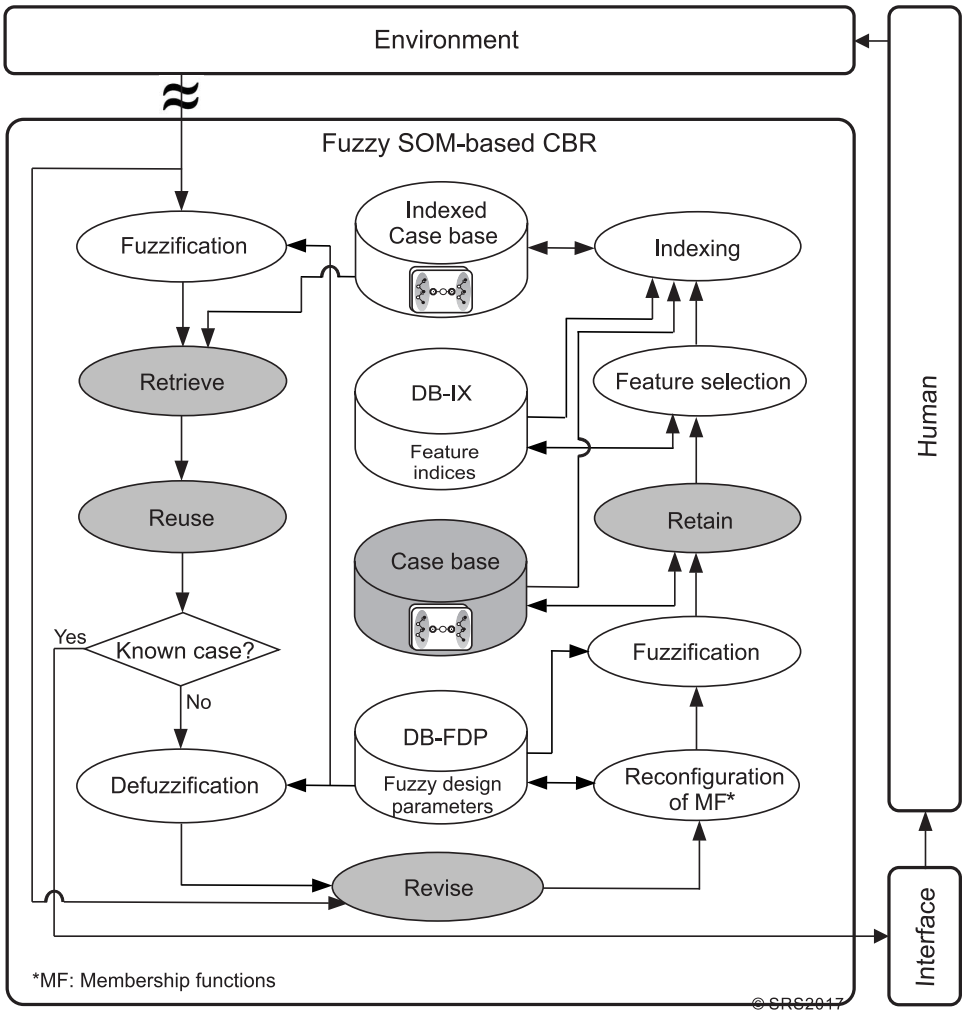

Figure 4.1: Fuzzy SOM-based CBR cycle [SHS17b]

base consists of $CaseBase$, $CaseBase_{fuzzy}$, $CaseBase_{indexed}$, DB_{IX}, and DB_{FDP} which are alternatively representing the original case base, fuzzy case base, indexed case base, database of indexing information, and database of membership functions design parameters.

Situation recognition procedure is performed repeatedly over time. This procedure is only processed if an event occurs. An event occurs if a new situation compared to the previously generated situation S_{i} happens. This comparison is done based on the fuzzy transformed values of the characteristics.

In this occurrence, the actual situation is considered as S_{i}. The situation S_{i} and active operator O are added to the sequence list of situations and operators. The procedure progresses to recognize the actual situation S_{i} by calling the retrieve

Algorithm 1

```
1: procedure SITUATION-RECOGNITION(t,S_i,O,List_sequence)
2:     //t : actual time point
3:     //S_i : last occurring situation
4:     //O : active action
5:     //List_sequence : a sub sequence of occurring situations and operators
6:     DataVector ← perception(t);
7:     S_t ← fuzzification(DataVector, DB_FDP);
8:     if S_t <> S_i then
9:         S_{i-1} ← S_i;
10:        S_i ← S_t;
11:        List_sequence ← add(List_sequence, O, S_i);
12:        SimilarCases ← retrieve(CaseBase_indexed, DB_IX, S_{i-1}, O, S_i);
13:        if SimilarCases <> null then
14:            InferredS_i ← reuse(SimilarCases, S_{i-1}, O, S_i);
15:            if Similar(S_i, InferredS_i) > τ then
16:                Pattern_i = identification(SimilarCases, S_i);
17:                return Pattern_i;
18:            end if
19:        end if
20:        cS_{i-1}, cO, cS_i ← defuzzification(S_{i-1}, O, S_i, DB_FDP);
21:        Case ← revise(cS_{i-1}, cO, cS_i);
22:        learning(Case, List_sequence);
23:    end if
24: end procedure.
```

procedure to find similar cases to the sequence of $S_{\mathrm{i}-1}$, O, and S_{i}. If there are some similar experiences to the considered event, they would be reused and adapted with the actual event through a fuzzy inference process. If the similarity between S_{i} and the result of case inference is acceptable (more than threshold τ_{s}, where $\tau_{\mathrm{s}} \in [0, 1]$ is defined by human experts), the experience is valid for the actual situation. Here, the situation pattern could be returned as a valid pattern. A new event could be detected if no similar cases are found or the inferred situation is not adapted to the actual situation. If the actual situation is new and needs to be saved in the case base, the procedure will be continued to revise and learn the new event (as explained in the next section).

Defuzzification of the sequence $S_{\mathrm{i}-1}$, O, and S_{i} is required before the revise procedure to transform the characteristics values to real scalar and precise values for case generation. Through the revise procedure and using the input sequence, a new crisp case $Case$ is generated. The learning process discussed in the next section, is consequently called to maintain the related knowledge base for further situation recognition.

4.2.1 Real-time situation recognition

Actual situation defined by a set of quantitative characteristics is considered as input of recognition process. The characteristics with crisp values are fuzzified based on the membership functions generated using learning process. Using a trained classifier in learning process, the situation pattern could be recognized and informed to human operators.

4.2.2 Similarity assessment

Using the indexed case base and through a similarity assessment algorithm in retrieve process, the most similar experienced cases are recalled. A local similarity assessment algorithm is applied to compare the simple characteristics the actual situation and experienced situations. A global similarity assessment algorithm is then used to compare compound characteristics of the situations. Therefore, the first two levels of situation recognition are accomplished through retrieve process. In this level of recognition process, the possible operators (actions) and the upcoming situations are identified. If the similarity assessment algorithm does not find any qualified experienced situation, the actual situation is identified for a new case, else the reuse process will be accomplished with further measurements for identification of new cases. Similarity assessment algorithms are designed based on the case representation approach to take comparable characteristics of a case into account.

4.3 Learning

The learning procedure presented in Algorithm 2 will assign situation patterns to the situations (through a labeling procedure), update and maintain the case base and its related knowledge base with new experiences. This procedure starts with reconfiguration of membership functions and updating the database DB_{FDP} based on the new experience. The related procedure $reconfigurationMF$ applies an algorithm for automatic reconfiguration of functions based on a density-based clustering method. Accordingly, the new case is fuzzified based on the reconfigured functions.

In the next step, the situations listed in $List_{sequence}$ are labeled/re-labeled for further situations recognitions if a specific situation pattern happens. Afterward, the $List_{sequence}$ would be started from the actual situation for the next labeling procedure. Labeling procedure should be configured separately for each application (see the labeling procedure written in Section 5.3.3 for lane-change situation recognition in driving maneuvers).

Retain procedure is applied for updating the original and fuzzy case bases using the new cases. Case reduction and feature selection considered as data reduction process are important tasks of case base maintenance. Case reduction is proceed through the retain procedure. Feature reduction is applied by a feature selection procedure to define the most relevant characteristics of the situations representing the situation patterns. The vector of measured importances for the features is captured in the database DB_{IX} for case base indexing procedure. Finally, the fuzzy case base could be indexed according to the most relevant characteristics.

Algorithm 2

1: **procedure** LEARNING($Case, List_{sequence}$)
2: $//Case : new\ case\ in\ crisp$
3: $//List_{sequence} : a\ sub\ sequence\ of\ occurring\ situations\ and\ operators$
4: $DB_{FDP} \leftarrow reconfigurationMF(DB_{FDP}, S_i)$;
5: $Case_{fuzzy} \leftarrow fuzzification(DB_{FDP}, Case)$;
6: **if** $definedpattern(Case_{fuzzy})$ **then**
7: $labeling(List_{sequence})$;
8: $List_{sequence} \leftarrow S_i$;
9: **end if**
10: $CaseBase \leftarrow retain(CaseBase, Case)$;
11: $CaseBase_{fuzzy} \leftarrow retain(CaseBase_{fuzzy}, Case_{fuzzy})$;
12: $DB_{IX} \leftarrow featureselection(CaseBase_{fuzzy}, DB_{IX}, Case_{fuzzy})$;
13: $CaseBase_{indexed} \leftarrow Indexing(CaseBase_{fuzzy}, DB_{IX})$;
14: **end procedure**.

4.3.1 Identification of new cases

Through the reuse process, a fuzzy inference is performed on the retrieved cases in the form of Eq. 3.3 and the actual situation S_i. The output expresses an adapted case with the actual situation S_i. The output is consequently defuzzified to determine the result to a corresponding crisp values to be comparable with the real-world upcoming situation. By revise process, the adapted case including possible operator and estimated situation is compared with the action performed by human operator and the upcoming situation alternatively. If the occurring situation is not similar to the next situation in the adapted case, it is marked as a new case. Modeling the sequence of situations using SOM approach could accomplished the third level of recognition through reuse process.

4.3.2 Learning the new cases

Learning new cases and updating the case base for further recognitions build one option to realize the learning process. Learning process in the proposed CBR is realized through the following sub-sections.

Labeling the cases

If a case is identified as a new case, it is learned as new experiences through a learning process. Here, the occurring situations could not be labeled in advance or immediately. Labeling the situations during online learning is only possible when specific situations (goal situations) occur. Here, SOM approach is used to generate a sequence of situations and performed operators to meet a goal situation, where the first unlabeled situation is defined as an initial situation for the sequence. By reaching to a goal situation, the occurring situations in the sequence are labeled as that goal situation. Finally, the approved cases related to the occurring situations are labeled and applied to the case base.

Case base maintenance

Case base maintenance which is widely studied in machine learning science, is a process for reorganization of the case base to make them reusable for further retrieval. This process goes forward with different tasks for updating the case base, revising the case representation and reconfiguration of related information. According to [LJP14], the goal is to obtain a reduced case base with the same problem-solving competence to be used as an index to enhance the retrieval of similar cases.

Fuzzification and generalization of comparable cases could be considered as an effective solution in case base reduction [BTRC+16]. Reconfiguration of fuzzy membership functions is assumed as a task of case base maintenance in fuzzy SOM-based CBR. It attempts to simplify the applicability of new cases with individualized linguistic terms. However, fuzzification of the continues environment variables with the target of individualization is a challenge as expert knowledge in definition of membership functions is not always available and reliable. Automatic reconfiguration of membership functions (MFs) is an important step of learning to modify design parameters of membership functions. Design parameters are investigated and probably regenerated based on the new distributions of the values for each quantitative characteristic. In this work, a density-based classification method [DHO08] is applied for automatic generation of membership functions. New parameters of fuzzy membership functions are updated in the database of fuzzy design parameters DB_{FDP}. Hereby, membership functions are defined and stored individually for humans operators. New cases are fuzzified using those membership functions.

Retain is an important task of case base maintenance. Through the retain process, the case base could be updated by adding new cases. The fuzzy cases are stored in an indexed case base for redundancy removal. As the feature values of the fuzzy data are assigned to one of the membership functions, multiple cases may be identical. As a result of this, the cases with the same fuzzy values could be considered as redundant knowledge. Therefore, a data reduction process is applied to reduce the redundancy. Data reduction correlates with the degree of generalization achieved by assigning data to the membership functions. Finally, through an indexing task, the case indices are regenerated with the new cases.

4.4 Generation of membership functions

Fuzzification has a big contribution to deal with uncertain knowledge through utilization of fuzzy sets to make them appropriate for approximate reasoning [dBBL17]. Fuzzy sets are defined using membership functions expressing linguistic variables. The membership degrees between 0 and 1 are assigned to each quantitative value to represent them using linguistic variables. The result of a fuzzification process is a vector containing the membership degrees of an element to all linguistic variables.

Generation of fuzzy membership functions is an essential part of performing fuzzy-based solutions. Using human expert knowledge is a basic method for definition of membership functions. However expert knowledge is not always reliable or available (and may possess a subjective character) specially in handling complex data. Therefore, designing suitable methods for automatic generation of fuzzy membership functions would be helpful. The corresponding methods should represent data correctly in fuzzy form by reducing dependence on expert knowledge.

Different methods have been formerly proposed for automatic generation of membership functions, such clustering-based [MAMA14, NU09], density-based [HO07, DHO08], heuristic [HL96], and data driven methods [WM14]. Density-based methods with the advantages of noise resistant, low computational time, and high clustering accuracy [HO07, DHO08] are the focus of this work. Density-based methods could be proposed by applying different algorithms such as CLUSTERDB*, FN-DBSCAN, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), Fuzzy Joint Points (FJP), and Noise-Robust FJP (NRFJP). According to [HO07, NU09], CLUSTERDB* and FN-DBSCAN are two common methods applied for density-based method. The approach CLUSTERDB* is able to determine the number of membership functions automatically without the need of predefined parameters [DHO08]. According to [NU09], the density-based method with FN-DBSCAN shows high accuracy and fast computational time. However, FN-DBSCAN needs a predetermination of some setting parameters. To compensate this drawback, a solution to automatically determine these parameters could be added to the algorithm. Additionally the authors showed that FN-DBSCAN algorithm provides more robust results than crisp DBSCAN algorithm, while running faster than other fuzzy neighborhood relation-based algorithms such as FJP [RFS16] and NRFJP [JSK16, YU16]. In this work, CLUSTERDB* and FN-DBSCAN are chosen for development of membership functions for continues variables.

Main contribution of this section is development and evaluation of the qualified methods for generation of membership functions to improve pattern recognition in the domains in which information is incomplete or imprecise.

4.4.1 Membership functions

Linguistic values are subsets of their respective linguistic variable and are represented by membership functions. According to [SBK16, MW16], trapezoidal membership function is often applied for different applications

$$\mu(x) = \begin{cases} 0 & \text{if } x < x_a, \\ \dfrac{x - x_a}{x_l - x_a} & \text{if } x_a \leq x \leq x_l, \\ 1 & \text{if } x_l \leq x \leq x_r, \\ \dfrac{x_e - x}{x_e - x_r} & \text{if } x_r \leq x \leq x_e, \\ 0 & \text{if } x > x_e, \end{cases} \tag{4.1}$$

where x_l and x_r describe the left and right borders of trapezoids core. The parameters x_a and x_e describe the beginning and end of the support for both functions respectively. A crisp value could be expressed in terms of several related linguistic variables. Values of those variables are measured using membership degree of a crisp

value to the respective functions. Accordingly, the crisp value x could be established by its corresponding fuzzy value $x_f(x)$ using the membership vector

$$x_f(x) = [\mu_{wL1}(x), \mu_{wL2}(x), ..., \mu_{wLn}(x)], \tag{4.2}$$

whose elements represent the membership degree to the membership functions $\mu_{wLi}(x)$ of the linguistic values wLi. Here, $1 \leq i \leq n$ and n indicates the number of membership functions. In the sequel, an overview of existing density-based methods for determination of membership functions design parameters is given.

4.4.2 Density-based methods

An approach for determination of design parameters could be given by using a clustering algorithm in combination with a density function. The process of membership functions generation using density-based methods could be shown as Fig. 4.2 and described shortly according to [AWS14, RL14] in this section.

Data clustering is the first step of this process. Given a set of clustered data C, the centroid of each cluster could be determined. The centroid is a data point which possesses the mean value of all data points within the cluster. If no such data point is found, the centroid is defined by finding the data point with the closest value to the mean. Once the centroids are found, the neighbors of the centroid should be detected.

The density of each neighbor in the set of centroid neighbors is measured by comparing its density to a threshold t. If the density of a neighbor is greater than or equal the threshold, the neighbor is considered as a dense neighbor. Its own neighbors which are not yet included in the set of centroid neighbors are added to the set. When the density of all the neighbors in the set is examined and no more dense neighbor is found, the set of membership function cores is completed. This process is repeated until all the clusters are processed. For each set of neighbors, the core and support parameters of each membership function (MF) are defined and returned as design parameters. In the following sub-sections, two density-based methods using CLUSTERDB* and FN-DBSCAN algorithms are adapted and implemented for generation of membership functions.

CLUSTERDB*

An approach for automatic generation of membership functions is given by applying CLUSTERDB* algorithm in combination with a density function. As described in [HO07], CLUSTERDB* algorithm is an improved version of the CLUSTER algorithm by using a validity index $DB*$ to prevent the generation of small and not well

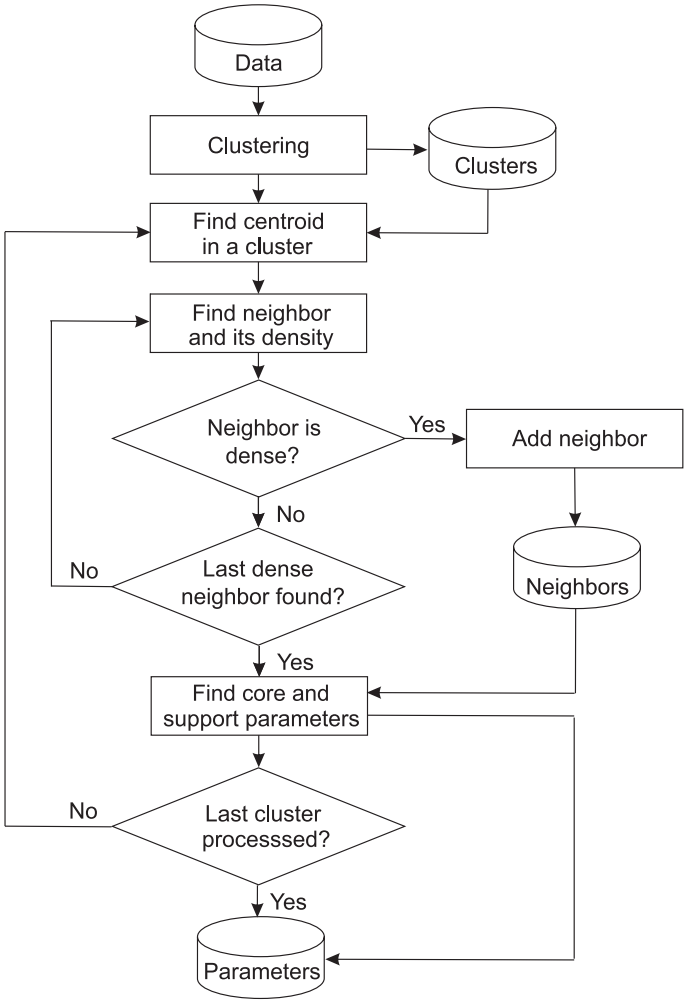

Figure 4.2: Membership function generation using density-based methods [SHNSon]

separated clusters. The DB^* variable describing the similarity of whole clusters is defined as [HO07]

$$DB^*(nc) = \frac{1}{nc}\sum_{i=1}^{nc}\left(\frac{max_{k=1,..,nc,k\neq i}\{S_i + S_k\}}{min_{p=1,..,nc,p\neq i}\{d_{ip}\}}\right), \tag{4.3}$$

where nc is the number of clusters, d_{ip} is the distance between the i-th and p-th centroids, and S_i and S_k address the scatter distances of the i-th and k-th clusters alternatively. In each iteration of CLUSTERDB* algorithm a relative neighborhood graph (RNG) is generated. Here, the adjacent data are connected with edges [HDO10]. The edges longer than a defined threshold β are removed for generation of isolated clusters. Finally the DB^* index is recalculated. The above steps are repeated for each cluster unless the stop conditions described in [HO07] are satisfied. As mentioned in [NU09], this algorithm was used on multiple data sets and has

shown to automatically generate an appropriate number of membership functions as well as determination of the design parameters with large cores.

Here, the methods are used to generate membership functions for each variable of a data set individually. Therefore, the generation of a 1-dimensional RNG is required for each variable since data points in the 1-dimensional feature space only have a maximum of 2 neighbors. This can quickly lead to over-partitioning of the data even in dense areas of the feature space. Hence, the presented CLUSTERDB* algorithm is adapted by applying the approach presented in [HL96] for generation of a 1-dimensional RNG. At first, the similarity value s is calculated for all adjacent data points. Adjacent points can be connected by edges if their similarity value is larger than threshold α. To prevent over-partitioning, a small value should be chosen for α.

Another adaption is related to the removal of the termination criteria which ends the algorithm if the number of clusters exceeds the square root of the data set. Here, the CLUSTERDB* algorithm is implemented with the following steps and is terminated when all the clusters are processed.

1. Define α and C.
2. Generate RNG.
3. Find centroids.
4. Calculate the DB^* using Eq. 3.6.
5. Set $i = 1$.
6. Calculate edge lengths for cluster C_i and sort them in ascending order.
7. Calculate variations of successive lengths and the intermediate variation as proposed in [HO07].
8. Find threshold β for C_i by fulfilling the criteria explained in [HO07]. If no threshold is found, set $i = i + 1$ and repeat from step 6.
9. Find edge greater than t. If no edge is found, set $i = i + 1$ and repeat from step 6.
10. Remove the edge and recalculate DB^*. If the resulting two clusters increase DB^*, undo the removal and recalculation of DB^* and repeat from step 8. If DB^* is not increased, repeat from step 6.
11. Terminate if all clusters are processed.

To get the optimal number and values of design parameters of the membership functions, an optimal value of threshold α should be defined. The value of parameter C does not have to be optimal, as a change in the value of C changes the similarity values s with the same magnitude [HL96]. In this work, genetic algorithm [Gre86] is applied for estimation of optimal value of α.

For implementation of the genetic algorithm defining the optimal value of α, the operations of roulette wheel selection, uniform mutation, and two-point crossover

have been utilized. The chromosomes are comprised of the α values for each variable of a data set. To indicate the fitness of a chromosome, the membership functions resulting from the chromosome are used to fuzzify a data set. The fuzzify data is then used to train and test of a classifier. The classification accuracy of the test data is used as fitness value of the chromosome.

Fuzzy Neighborhood Density-Based Spatial Clustering of Applications with Noise (FN-DBSCAN)

Another clustering algorithm namely FN-DBSCAN [NU09, CMS+10] has been proposed to improve the CLUSTERDB* algorithm. The FN-DBSCAN algorithm assigns fuzzy neighborhood values representing the degree with which a data point y belongs to the neighborhood of another data point x. Based on the neighborhoods, FN-DBSCAN searches for fuzzy core points to generate clusters. According to [NU09], by considering N_x as a neighborhood membership function of the point $x \in X$, point x is considered as a fuzzy core, if its neighborhood

$$N(x) = \{y \in X \mid N_x(y) \geq \epsilon_1\}, \tag{4.4}$$

with a membership degree greater than a predefined threshold ϵ_1 possesses a fuzzy cardinality

$$cardFN(x) \equiv \sum_{y \in N(x)} N_x(y) \geq \epsilon_2, \tag{4.5}$$

more than the predefined threshold ϵ_2. In [NU09], the FN-DBSCAN algorithm has been proven to detect efficient number of clusters depending on the choice of ϵ_1, and ϵ_2.

Here, FN-DBSCAN algorithm is adapted and implemented to cluster the data for density-based generation of trapezoidal membership functions. As suggested in [NU09], the data should be initially normalized and sorted in ascending order before applying the FN-DBSCAN algorithm. Once the data are normalized, the parameters ϵ_1, and ϵ_2 defined in Eq. 4.4 and Eq. 4.5 need to be found. To determine each data point fuzzy neighborhood, the linear membership function presented in [SHNSon] is used.

After the data are clustered and the centroids are defined, the membership functions are generated by determining the cores through calculating each data points density value [NU09]. Based on the density values, the dense neighbors of the centroids are determined as addressed in [DHO08]. The centroids and their neighbors together build the core. Once the cores are generated, the supports are determined as ranging from the right core boundary of the previous core to the left boundary of the subsequent core.

The application of FN-DBSCAN requires expert knowledge to a certain degree, since the parameters ϵ_1, and ϵ_2 need to be predefined. As explained in [NU09], the average distance between adjacent data points is considered as the value of neighborhood threshold. Every data point y which is closer to the data point x than the average distance of adjacent data, possesses a neighborhood degree of $N_x(y) > 0$ [NU09].

The proposed FN-DBSCAN algorithm could be adapted to work more independence from experts knowledge by setting the parameter ϵ_1 to zero. The parameter ϵ_1 describes a radius, defining the membership threshold for data points to be considered in the fuzzy cardinality. By setting ϵ_1 to a value larger than zero, the density requirements towards the center of the neighborhood could increases. Accordingly, this is achievable by setting the fuzzy cardinality threshold ϵ_2 to avoid the necessary effort for determination of ϵ_1. Accordingly, the FN-DBSCAN algorithm could be implemented as follows

1. Define ϵ_2 and C.
2. Normalize and sort data.
3. Calculate the mean distance ϵ between adjacent data.
4. Set $i = 1$.
5. Find an unassigned fuzzy core point p for cluster C_i fulfilling the criteria given in Eq. 4.4 and 4.5. If no unassigned fuzzy core point is found, terminate the algorithm.
6. Assign p to the cluster C_i.
7. Add all unassigned neighbors of p to an empty set of seeds S.
8. Select a point q from S and assign q to C_i.
9. If q is a fuzzy core point fulfilling the same criteria as in step 5, add the unassigned neighbors of q to S.
10. Repeat steps 8 and 9 until all seeds are assigned.
11. Set $i = i + 1$ and repeat from step 5.

In this algorithm, the number of predefined parameters is reduced from 3 to 1. The only parameter to be determined is the fuzzy cardinality threshold ϵ_2. In this work, a genetic algorithm is applied for estimation of ϵ_2 using the operations of roulette wheel selection, uniform mutation and two-point crossover. Here, chromosomes contain the ϵ_2 value of each variable of a data set.

4.4.3 Evaluation process

To evaluate the performance of the presented methods, classification performance based on the crisp and fuzzy data after fuzzification process is measured as shown in Fig. 4.3. In the classification process shown in Fig. 4.3, a data set of crisp values is considered as input and the output is a classified data set. The input data could

be classified in two types, (I) as fuzzy data in which input data should be fuzzified using the proposed approaches or (II) as crisp data.

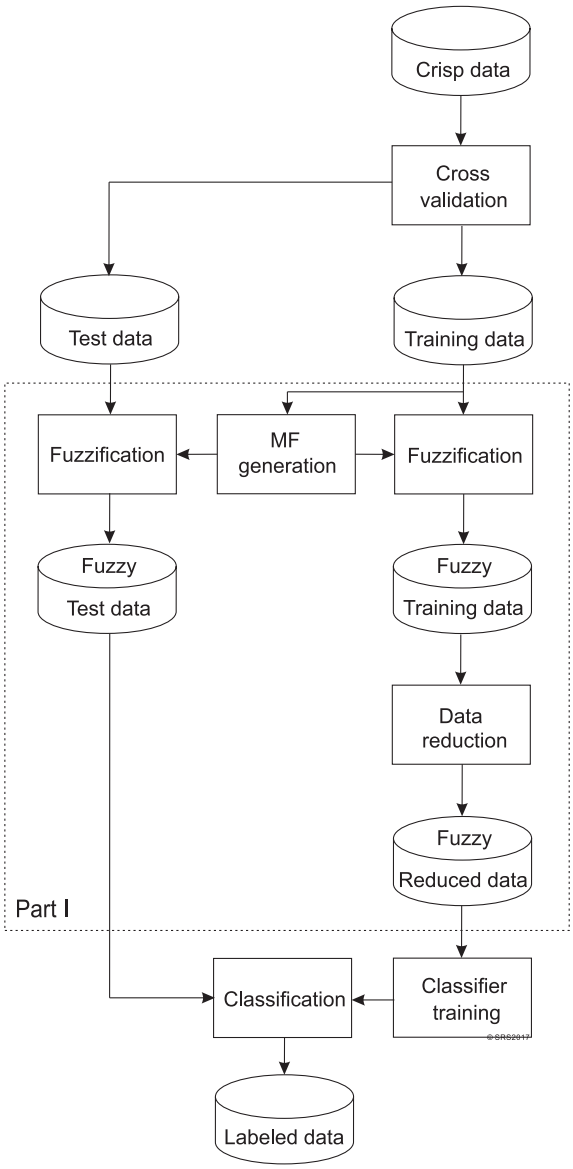

Figure 4.3: Classification with crisp and fuzzy data; part I is only performed for classification of fuzzy data [SHNSon]

In both classification processes I and II, the data set is separated into training and test data. To validate the generated membership functions, 10-fold cross-validation is used. Training data are applied for generation of a classifier and test data applies for evaluation of the classifier by classification of unknown data.

In the process I, the training data should be applied for automatic generation of membership functions by using one of the implemented algorithms before making a classifier. By means of the generated membership functions, the test and training data are fuzzified differently. To fuzzify the test data, the feature value of each instance is assigned to the membership function returning the greatest membership degree. On the other hand, the feature value of each instance in training data is assigned to the membership functions returning membership degrees greater than 0.4.

A training data instance with a feature assigned to multiple membership functions split into multiple instances. The instances contain the single different cases of that feature, while the assignments for the rest of the features are duplicated. The training data may therefore have more instances after fuzzification.

As the feature values of the fuzzy data are integers representing the assignment to one of the membership functions, multiple instances of the fuzzy data may be identical. Therefore, a process of data reduction is applied to reduce the fuzzy training data. After the data reduction, only the uniquely occurring instances in respect to their feature assignments and class affiliation are left. The classifier is trained using the reduced data. The trained classifier is then tested by means of the fuzzy test data.

4.5 Feature selection in fuzzy SOM-based CBR

Feature selection is the process of defining a minimal subset of available characteristics which satisfying a given criteria. Here, irrelevant and redundant characteristics with little or no predictive information about the defined situation patterns are eliminated. Feature selection consists of several steps. At first, using a proper algorithm the relevant characteristics for special patterns are defined and weighted, then the characteristics are ranked. Finally, using a selection strategy, the best set of characteristics is chosen and stored in DB_{IX}.

Feature selection is known as an effective process to improve the ability of situation recognition [ZHQ+15]. It allows to choose a set of relevant characteristics recognizing predefined patterns of situations.

Different algorithms have been proposed by previous studies for feature selection in different applications. Feature selection algorithms are divided into two main classes: wrappers-based and filter-based. Wrapper-based algorithms estimate the performance of the algorithm after each sub-set selection and have a high computational complexity as they apply a learning algorithm continuously [PF15, VR14]. Filter-based algorithms discard irrelevant features before learning, so they are faster than wrapper algorithms [PF15]. Rough set theory (RST) [Paw12] used for measuring the data dependencies and dimensionality reduction is applied widely for

filter-based algorithms [JS07], [TP09], [AE13] in real time applications [PA15]. It helps to find the key features contained in a database without additional information. Different generic filter-based algorithms have been proposed based on RST and evaluation functions for calculating the suitability of selected features [AE13]. The Quick Reduction (QR) algorithm [JS02] widely applied by previous studies calculate a reduct without generating all possible subsets [AE13]. However, QR can not guarantee to find the optimal sub-set of relevant features. In [SLS11], RST is applied for measuring the feature relevance. Three feature selection strategies are proposed for finding a sub-set of relevant features. These strategies depend on the selection criteria: Mean Selection (MS), Half Selection (HS), and Selection by Threshold (ST) based on RST. According to the results presented in [SLS11], feature selection using the proposed strategies overcome some problems related to the standard feature selection algorithms such as JohnsonsReduct applying a heuristic function and decision Table [GAM06]. In this section by focusing on feature selection, the filtering algorithm and three selection strategies presented by [SLS11] are applied to the process. Then, the performance of situation recognition process based on different selected sub-sets is evaluated and compared for an experimental application on driving maneuvers.

4.5.1 Feature Selection using RST

Rough set theory mainly deals with classification analysis of data sets that contains vagueness, missing values, or data redundancy [SGS14]. Using this theory, it is possible to reduce data redundancy and define common patterns in uncertain and incomplete data sets. The main advantage of RST is that only the structure of data and no more information about the membership values are required to define a suitable subset of relevant features.

Rough set theory is defined here. Let $I = (U, A)$ be an information system, where U is a non-empty finite set of objects and A is a non-empty finite set of characteristics that $\forall a \in A,\ a : U \rightarrow V_a$ and V_a are the possible values of characteristic a. For $B \subseteq A$ an associated equivalent relation R_B

$$R_B = \{(x, y) \in A^2 | \forall a \in B, a(x) = a(y)\} \quad , \tag{4.6}$$

exists. Let $X \subseteq U$. Set X is approximated only using the information contained within R_B. The lower and upper approximation of set X is given as

$$R_B \downarrow X = \left\{x \in U | \left[x\right]_{R_B} \subseteq X\right\} \quad , \tag{4.7}$$

$$R_B \uparrow X = \left\{ x \in U \,|\, [x]_{R_B} \cap X \neq \emptyset \right\} \quad , \tag{4.8}$$

where $[x]_{R_B}$ denotes equivalent classes of R_B including x.

Rough Set is defined using the tuple $\langle R_B \downarrow X, R_B \uparrow X \rangle$.

Positive region $\mathrm{POS}_B(X)$ calculates the lower approximation for each decision features and gives the union of the calculated approximations as

$$POS_B(X) = \bigcup_{\forall x \in X} R_B \downarrow x \quad . \tag{4.9}$$

4.5.2 Measurement of features relevance

To measure the importance of X by mapping the objects into R_B, the degree of dependency $\gamma_{\mathrm{B}}(\mathrm{X})$ $(0 \leq \gamma_{\mathrm{B}}(\mathrm{X}) \leq 1)$ [Kun92] is calculated as

$$\gamma_{R_B}(X) = \frac{|POS_B(X)|}{|U|} \quad . \tag{4.10}$$

The value $\gamma_{R_B}(X) = 1.0$ denotes that X depends only on R_B.

Relevance of each feature could be measured using the degree of dependency to indicate how the classifier changes by discarding the feature. The relevance of feature f, $\sigma(f)$ is calculated as

$$\sigma(f) = \gamma_{R_B}(D) - \gamma_{R_B - \{f\}}(D) \quad , \tag{4.11}$$

where R corresponds to the smallest possible reduction in decision table.

Algorithm 3 Quick reduction algorithm inspired from [JS04]

```
1: procedure QuickReduct(C,D)
2:     //C : set of all conditional characteristics;
3:     //D : set of all decision characteristics.
4:     R ← ∅;
5:     do
6:         T ← R;
7:         for each x ∈ (C − R) do
8:             if γ_{R∪{x}}(D) = γ_T(D) then
9:                 T = R ∪ {x};
10:            end if
11:        end for
12:        R ← T;
13:    while γ_R(D)! = γ_C(D)
14:    return R;
15: end procedure.
```

4.5.3 Quick reduction algorithm

The function illustrated by Algorithm 3 is a QR algorithm determining a set of relevant characteristics that is sufficient to specify the decision characteristics. This function attempts to choose a subset of conditional characteristics (without generating all possible subsets) as relevant characteristics for solving a defined problem. However, the output may be a close-to-minimal subset of characteristics and is not guaranteed to find the best combination of relevant characteristics.

Here, QR algorithm could be applied for feature selection to the learning process. The set of cases in CBR is considered as set U. Available characteristics defining a situation are shown as the set of conditional characteristics C. Available situation patterns are denoted as decision characteristics D. The reduction algorithm returns a reduced set of important characteristics of a situation for recognizing specified situation patterns.

4.5.4 Feature selection strategies

Feature selection as a filter-based algorithm is developed using three different selection strategies [SLS11] to determine the best relevant subset of features. Using the feature relevance measured by RST, a relevant sub-set is selected through a feature selection strategy.

Mean selection (MS)

According to this selection strategy, feature $f \in A$ is selected if

$$\sigma(f) \geq \sum_{a \in A} \frac{\sigma(a)}{|A|} \ , \tag{4.12}$$

where $\sigma(f)$ and $\sigma(a)$ are relevance values for the features $f, a in A$.

Half selection (HS)

According to this selection strategy, a feature is selected if its position in the rank vector is lower than approximately half of the features in the domain or the feature is an independent feature. Thus, feature $f \in A$ is selected if at least one the following conditions is satisfied

$$P_f \leq \lfloor \frac{|C|}{2} \rfloor \ , \tag{4.13}$$

$$\sigma(f) = 1.0 \ , \tag{4.14}$$

where P_f is the position of feature in the rank vector.

Selection by threshold (TS)

According to this selection strategy, a feature is selected if its relevance value is larger than the average relevance of the features in rank vector. Thus, $f \in A$ is selected if

$$\sigma(f) \geq min_r + \frac{max_r - min_r}{3} \ , \tag{4.15}$$

where $\mathrm{min_r}$ and $\mathrm{max_r}$ are minimum and maximum relevance values in the ranking vector, respectively.

Here, Algorithm 4 addresses the feature selection using RST applied to the proposed CBR. In the proposed algorithm, MS strategy is used for feature selection.

Algorithm 4 Feature selection using FRS theory and MS strategy inspired from [SLS11]

```
1: procedure FeatureSelection(C,D)
2:     //C : set of all conditional characteristics;
3:     //D : set of all decision characteristics.
4:     R ← ∅;
5:     for each x ∈ C do
6:         σ(x) = γ_R(D) − γ_{R−{x}}(D)
7:     end for
8:     for each x ∈ C do
9:         if σ(x) ≥ Σ_{a∈C} σ(a)/|C| then
10:            R = R ∪ {x};
11:        end ifend;
12:    end for
13:    return R;
14: end procedure.
```

4.5.5 Evaluation process

A feature selection framework consists of several steps shown in Fig. 4.4 are introduced for evaluation of the proposed feature selection algorithms.

Feature selection algorithms may not work successfully in real-time applications because of dealing with crisp values of the features [PA15]. By application of fuzzy theory and fuzzification process, the actual data is discretized and labeled with different degrees of membership. At first the features values are discretized by transferring into fuzzy values using the preconstructed fuzzy sets and predefined membership functions in DB_{FDP}. Then, the relevance of each characteristics for special patterns are measured. The characteristics are sorted by rank. Finally, using a selection strategy, the best set of characteristics is chosen and stored in DB_{IX}. The selected characteristics using different strategies are evaluated through the situation recognition process.

The proposed framework is used in the next chapter for evaluation based on the experimental data. Accordingly, an appropriate algorithm to be selected for integration with fuzzy SOM-based CBR approach.

4.6 Summary

In this chapter, the classical CBR approach is improved for individualized situation recognition to support the proposed knowledge representation. Accordingly, the

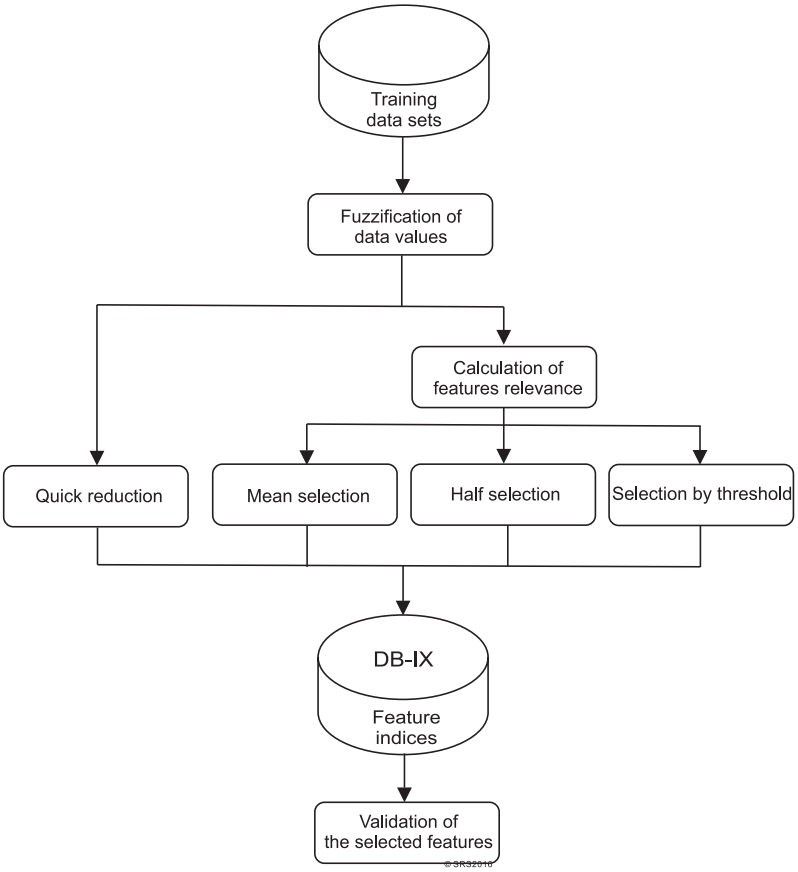

Figure 4.4: Feature selection framework [SS16b]

applied algorithms in realization of situation recognition as well as learning process to supervise individual behaviors of human operators are detailed. In the next chapter, the proposed framework is implemented and evaluated for driving situation recognition using the data acquired by a driving simulator.

5 Experimental results

In this chapter, the proposed Fuzzy SOM-based CBR approach is implemented for individualized recognition of lane-change situations applied to a simulated driving environment. In this application, lane-change situations could be recognized by considering and learning the exclusive lane-change behaviors of individual human drivers.

Situation recognition is evaluated using different test driving data related to 9 drivers with various lane-change behaviors. The evaluation results will state the ability of the proposed approach in real-time situation recognition for supervision of individual human operators in the applied application.

The contents, figures, and tables presented in this chapter have been prepared for publication as the journal papers "Fuzzy SOM-based Case-Based Reasoning for individualized situation recognition applied to supervision of human operators" [SHS17b] and "learning and representation of event-discrete situations for individualized situation recognition using fuzzy SOM-based CBR" [SHSon], and partly published as the conference paper [SHS17a].

5.1 Framework design for fuzzy SOM-based CBR

In this section, a framework is designed using a CBR-oriented software for implementation of the proposed fuzzy SOM-based CBR. The CBR-oriented software provides a reference platform for developing CBR applications and the related databases. The most common software used in recent years are CBR Shell [AIA17], FreeCBR [Lar17], myCBR [fAI17], eXiTCBR [oG17], and jCOLIBRI [BTGCDA04].

According to a comparative analysis of the software presented in [EM14, AA12], the advantages of jCOLIBRI justify the implementation of fuzzy SOM-based CBR in this work. Using this software, high-dimensional data and large case bases required for realization of Fuzzy SOM-based CBR could be managed.

Using the case structure ability in jCOLIBRI, the case class could be defined and applied successfully. Case indexing could only be supported by jCOLIBRI. Moreover, the evaluation of each software using the statistical measurements (precision, recall, and F-Measure) as well as accuracy shows that a CBR approach developed using jCOLIBRI outperforms the approaches developed using the other software.

5.1.1 Application of jCOLIBRI

This software is used for implementation of CBR platform for different complex applications. It is a Java-based framework and uses JavaBeans technology for case representation by supporting various types of databases including external databases.

The top layer of the COLIBRI platform is COLIBRI Studio. It contains visual tools for implementation of CBR cycle and automatic configuration of required libraries for generation of a CBR platform. It is integrated with Eclipse IDE to manage Java-based projects. Using COLIBRI Studio, the expert designer could define different templates where several tasks are linked together to specify the desired behaviors. The lower layer of jCOLIBRI includes a library of the classes for implementation of CBR system editor including previous cycle (PreCycle), main cycle (Cycle), and post cycle (PostCylce) of CBR as it is shown in Fig. 5.1. In PreCycle, the databases and case base connections are activated, and related temporary data sets are initialized. The main cycle supports the situation recognition and learning procedures. Finally, through the PostCylce, the data from temporary memory are stored into the related databases.

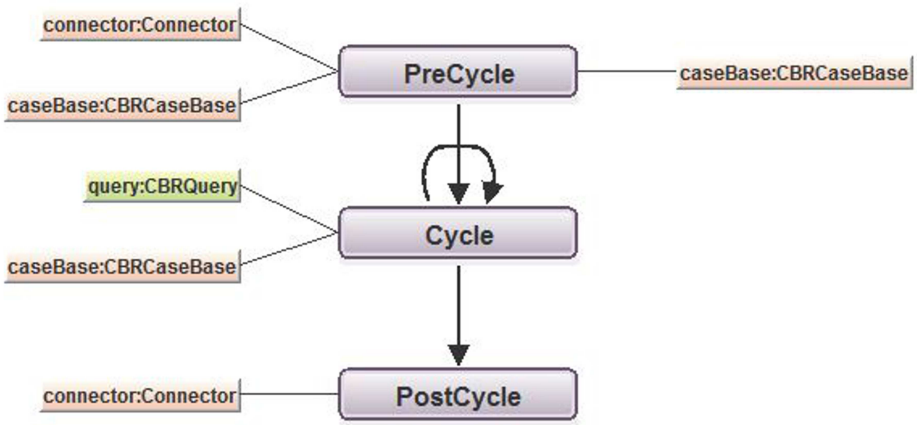

Figure 5.1: System Editor [using COLIBRI Studio]

5.1.2 Requirements and design overview

The proposed fuzzy SOM-based CBR approach is developed by definition of a CBR project using jCOLIBRI. Accordingly, case structure is configured using a case designer. A Case in jCOLIBRI is defined using a “description”, a “solution” and a “result” which contains a set of attributes (characteristics). There are two types of characteristics: simple or compound which is a function of simple and other

compound characteristics. Each simple characteristic is defined with a name, a data type, local similarity function, and its weight which is a value between 0 and 1. Each compound characteristic is defined using a name and a global similarity function. A part of configuration panel in jCOLIBRI is illustrated in Fig. 5.2. The terms "description", "solution", and "result" representing an initial situation, applied operator, and upcoming situation alternatively. For local similarity assessment, Nearest Neighbour scoring function is applied. In addition, the weights of the characteristics are initialized to 1.0.

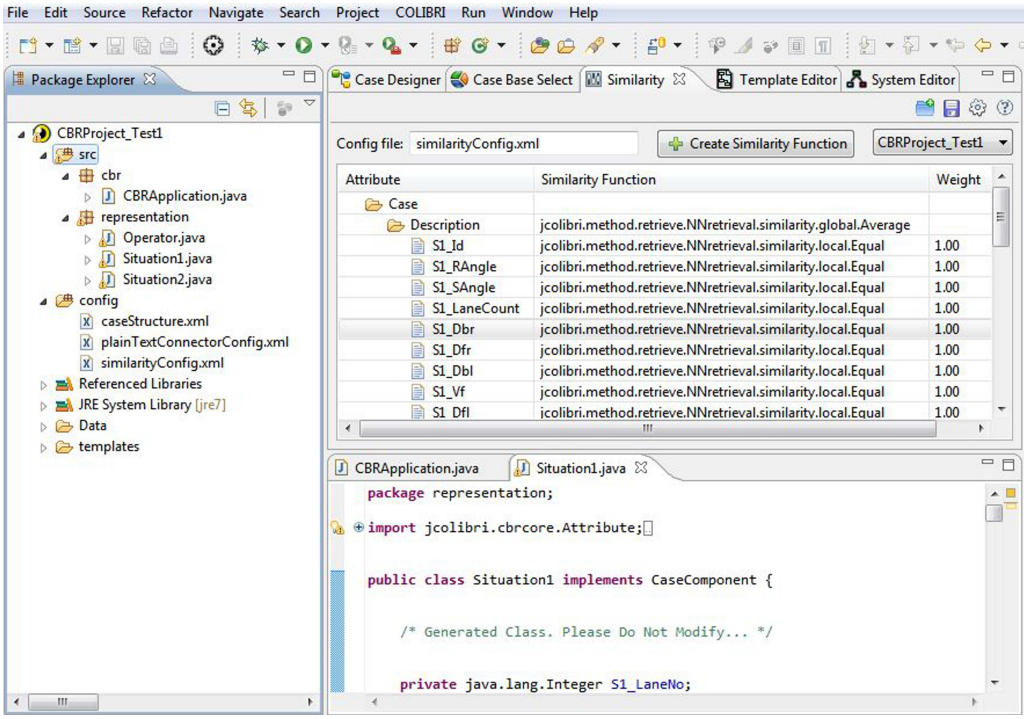

Figure 5.2: Configuration of case structure in jCOLIBRI

Several external libraries are applied to support the required functions and procedures for complementation of situation recognition (Algorithm 1) and learning (Algorithm 2) procedures. Those libraries are listed as in the following sub-sections.

Database connection

The case base and its related databases for storing the feature weights and membership functions design parameters, relational databases are configured and connected to the target project. Data for a CBR project in jCOLIBRI are configured in two layers of persistence mechanism and in-memory organization (for more information see [RSDAGC05]). Persistence data is supported with the concept of database

Connector. Connectors are objects to access the relational databases via the CBR project. The concept of connectors including the configuration of the required connectors in the target project is illustrated in Fig. 5.3.

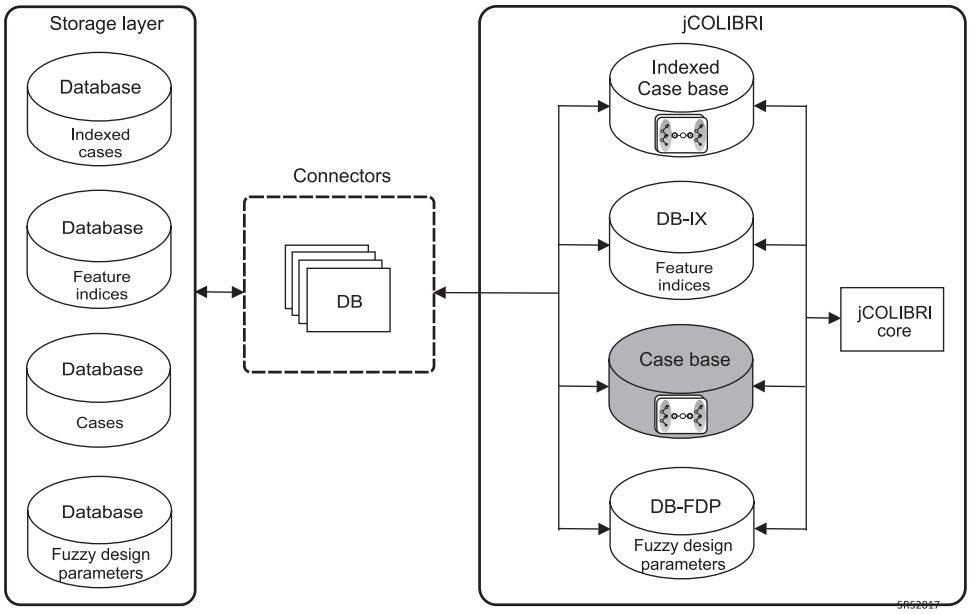

Figure 5.3: Database connectors architecture in jCOLIBRI

Supporting fuzzy logic

To design a fuzzy logic controller and inference engine, jFuzzyLogic library [CAF12] is applied to the target project implemented in jCOLIBRI. This library is an open source java library which supports fuzzification and defuzzification functions. Accordingly, fuzzy membership functions, fuzzy rules, as well as fuzzy inference engine are implemented and employed through this library.

Supporting learning methods

To apply the machine learning methods, k-NN and random forest, a java library called MULAN [TSXVV11] is applied to the target CBR project. This library offers different classification methods, feature selection, as well as dimensionality reduction algorithms.

5.2 Evaluation of fuzzy SOM-based CBR

In this section, the proposed situation recognition approach is developed and examined for recognizing driving lane-change situations. The performance of the proposed approach in situation recognition and learning is evaluated in different conditions and using various evaluation metrics.

5.2.1 Detection, false alarm, and accuracy rates

Situation recognition using the proposed approach is evaluated in terms of detection rate, false alarm rate, and accuracy of the recognition process. Here, detection rate, false alarm rate, as well as accuracy rate are measured with four possible outcomes as

True Positive (TP) Number of overtaking situations recognized correctly by the system,
False Negative (FN) Number of overtaking situations which are not recognized by the system,
True Negative (TN) Number of driving situations which are incorrectly recognized as overtaking situation by the system, and
False Positive (FP) Number of driving situations which are incorrectly recognized as overtaking situation.

Detection rate indicating the sensitivity of the recognition process is calculated as

$$Detection\ rate = \frac{TP}{TP+FN} \ , \tag{5.1}$$

A good recognition process should balance between the detection rate and the false alarm rate. Accordingly, False alarm rate is calculated as

$$False\ alarm\ rate = \frac{FP}{TN+FP} \ , \tag{5.2}$$

The effectiveness of situation recognition process could be evaluated using the recognition accuracy function as

$$Recognition\ accuracy = \frac{TP+TN}{TP+TN+FP+FN} \ . \tag{5.3}$$

5.2.2 Response time

Moreover, response time is an important metric to show the performance of the proposed approach for situation recognition and its learning process. Recognition elapsed time and learning elapsed time indicate the length of time required to perform situation recognition process and learning process alternatively. Consequently, they are considered as measures for online-applicability of the proposed approach.

Recognition elapsed time is calculated as

$$T_{Recognition} = T_F + T_S + T_U + T_I \ , \tag{5.4}$$

where T_{F}, T_{S}, T_{U}, and T_{I} are computational time required for fuzzification, retrieve, reuse, and identification processes alternatively.

In addition, learning elapsed time is calculated as

$$T_{Learning} = T_{MFG} + T_F + T_L + T_T + T_{FS} + T_{IX} \ , \tag{5.5}$$

where T_{MFG}, T_{F}, and T_{L} are computational time required for generation of membership functions, fuzzification, and labeling. The variable T_{T} addresses the time required for storing the cases and case reduction in fuzzy case base. Moreover, T_{FS} and T_{IX} are the elapsed times for feature selection and indexing alternatively.

5.2.3 Generalizability of approach

Generalizability of an approach with different combinations of lane-change maneuvers for training and test could be measured using n-Fold cross-validation technique. Here, to estimate the performance of the proposed situation recognition approach from available data, 10-fold cross-validation technique is applied. In this validation which is repeated 10 times, all the data eventually used for both training and test. This technique is needed to evaluate not only the recognition accuracy independent of the data but also generalizability of the proposed approach with different integrations of lane-change maneuvers for training and test. The generalizability is high when the discrimination for unknown data is not significant [Min08].

5.2.4 Knowledge base workload

The workload of knowledge base consists of a set of related databases could be measured for assessing the online-applicability of the learning process. A database workload is a set of transactions done against the database [NKM01, ACN02, BLNZ15]. Here, total workload of the knowledge base could be measured as

$$WL_{Total} = N_{CBRQueries} + N_{CBRUpdates} + N_{IXUpdates} + N_{FDPUpdates} \ , \tag{5.6}$$

where $N_{\text{CBRQueries}}$ determines the number of queries executed on the case base. In addition, $N_{\text{CBRUpdates}}$, $N_{\text{IXUpdates}}$, and $N_{\text{FDPUpdates}}$ address the number of update transactions done on the case base, DB_{IX}, and DB_{FDP} alternatively in a given period of time.

In this research, the normalized workload for knowledge base update is calculated in a given period of time during learning process for comparing the changes in recognition accuracy and update workload in that period. The workload for knowledge base updates is calculated as

$$WL_{Update} = N_{CBRUpdates} + N_{IXUpdates} + N_{FDPUpdates} \ . \tag{5.7}$$

Accordingly, the normalized workload for updates is calculated as

$$WL_N = \frac{WL_{Update} * 100}{WL_{Total}} \ . \tag{5.8}$$

5.3 Experiment for lane-change situation recognition

The purpose of driving assistance systems is to offer to drivers additional convenience and safety by supporting their driving task [BMS16]. In recent years, lane-change assistance system has been successfully proposed and may be applied to the vehicles in recent years. However, the proposed lane-change assistance systems can detect a potential hazard when the driver is changing lanes [BMS16].

The proposed lane-change situation recognition in this contribution would be able to recognize the situations potentially lead to a lane-change event without consideration of humans intention in changing the lane. The recognized situations could be applied for situation assessment to improve the situation awareness and decision-making of human drivers.

In this section, the proposed fuzzy SOM-based CBR approach is developed for individualized recognition of lane-change situations applied to a simulated driving environment. In this application, lane-change situations could be recognized by considering and learning from exclusive lane-change behaviors of a driver.

5.3.1 Experiment design

In this experiment, different test drivers have been invited to drive with a professional driving simulator located at the Chair of Dynamics and Control, University of Duisburg-Essen.

Using a designed scenario, the test drivers had to drive and experience different lane-change maneuvers to overtake other vehicles on a highway. When the ego-vehicle is faster than the vehicle driving in front and the same lane, the driver can change the driving lane to the left. Through a lane-change to the right, the driver can move back the vehicle to the right lane and continue forward driving. In this scenario, in average two lane-change maneuvers could be experienced per minute by a driver.

Here, situations are defined with a set of characteristics addressing the internal status of ego-vehicle and status of the vehicles in the surrounding area of the ego-vehicle compared to the ego-vehicle as shown in Table 5.1. Among these characteristics, steering wheel angle for ego-vehicle c_{26} and accelerator pedal c_{21} are considered for operator. In real-world traffic, however, some considerations are affecting lane-change behavior [LTB15] such as traffic signs or human drivers emotional states which are not considered in this work.

In this simulator study, the data are received and abstracted every 0.05 seconds. The test drivers drive with driving simulator several times. Therefore, different data sets related to each test driver are applied to the proposed approach for evaluation.

5.3.2 Situation patterns

In this application, three patterns should be experienced, defined, and recognized through the proposed approach: lane-change to left (LCL), lane-change to right (LCR), and lane-keeping (LK).

Four key points in each lane-change maneuvers are defined to specify LCL, LCR, or LK situations in a sequence of occurring events. The key points P_1, P_2, P_3, and P_4 shown in Fig. 5.4 are pointing to the approximate start point, middle point, exact point, and approximate end point of a lane-change maneuver alternatively. According to the specified points, different states representing different situation patterns are defined. The occurring situations between each two points along a lane-change maneuver belong to one of three states St_1, St_2, and St_3. The occurring situations in St_1 and St_2 are defined as lane-change to left/right patterns. The occurring situations which are not in St_1 and St_2, are specified as a lane-keeping pattern.

Table 5.1: Characteristics of a situation

Name	Characteristic	Range	Data type
c_1	Velocity of ego-vehicle [km/h]	[0 220]	Real
c_2	Lane number	{1,...,4}	Integer
c_3	Distance to vehicle in front [m]	[0 250]	Real
c_4	Distance to vehicle left-front [m]	[0 250]	Real
c_5	Distance to vehicle right-front [m]	[0 250]	Real
c_6	Distance to vehicle left-behind [m]	[0 250]	Real
c_7	Distance to vehicle right-behind [m]	[0 250]	Real
c_8	Distance to vehicle behind [m]	[0 250]	Real
c_9	Velocity of vehicle in front [km/h]	[0 220]	Real
c_{10}	Velocity of vehicle left-front [km/h]	[0 220]	Real
c_{11}	Velocity of vehicle right-front [km/h]	[0 220]	Real
c_{12}	Velocity of vehicle left-behind [km/h]	[0 220]	Real
c_{13}	Velocity of vehicle right-behind [km/h]	[0 220]	Real
c_{14}	Velocity of vehicle behind [km/h]	[0 220]	Real
c_{15}	TTC for vehicle in front [s]	[0 12]	Real
c_{16}	TTC for vehicle left-front [s]	[0 12]	Real
c_{17}	TTC for vehicle right-front [s]	[0 12]	Real
c_{18}	TTC for vehicle left-behind [s]	[0 12]	Real
c_{19}	TTC for vehicle right-behind [s]	[0 12]	Real
c_{20}	TTC for vehicle behind [s]	[0 12]	Real
c_{21}	Accelerator pedal	{1,...,3}	Integer
c_{22}	Break pedal	{1,...,3}	Integer
c_{23}	Gearbox	{1,...,4}	Integer
c_{24}	Indicator	{0,1}	Bit
c_{25}	Heading angle between ego-vehicle and road	[-3.14 +3.14]	Real
c_{26}	Steering wheel angle for ego-vehicle	[-3.14 +3.14]	Real

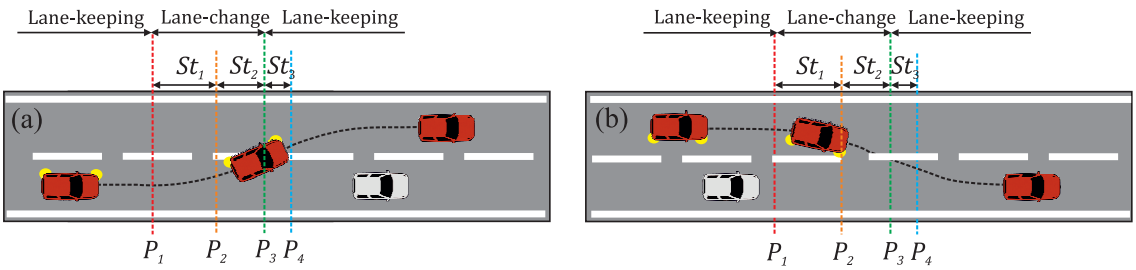

Figure 5.4: Different situation patterns along the driving maneuvers; (a) lane-change to left, and (b) lane-change to right

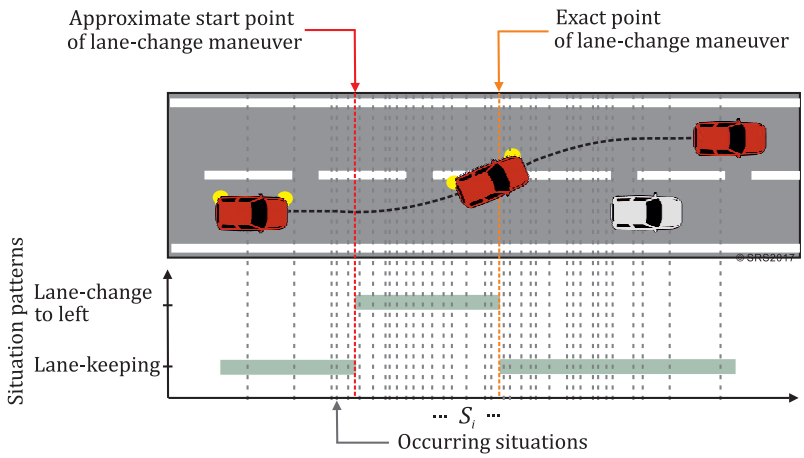

Figure 5.5: Occurring situations and identified situation patterns along a driving maneuver

5.3.3 Offline training, online training, and test

The experiment is organized into three phases for each human driver data: offline training, online training, and test (as shown in Fig. 5.6).

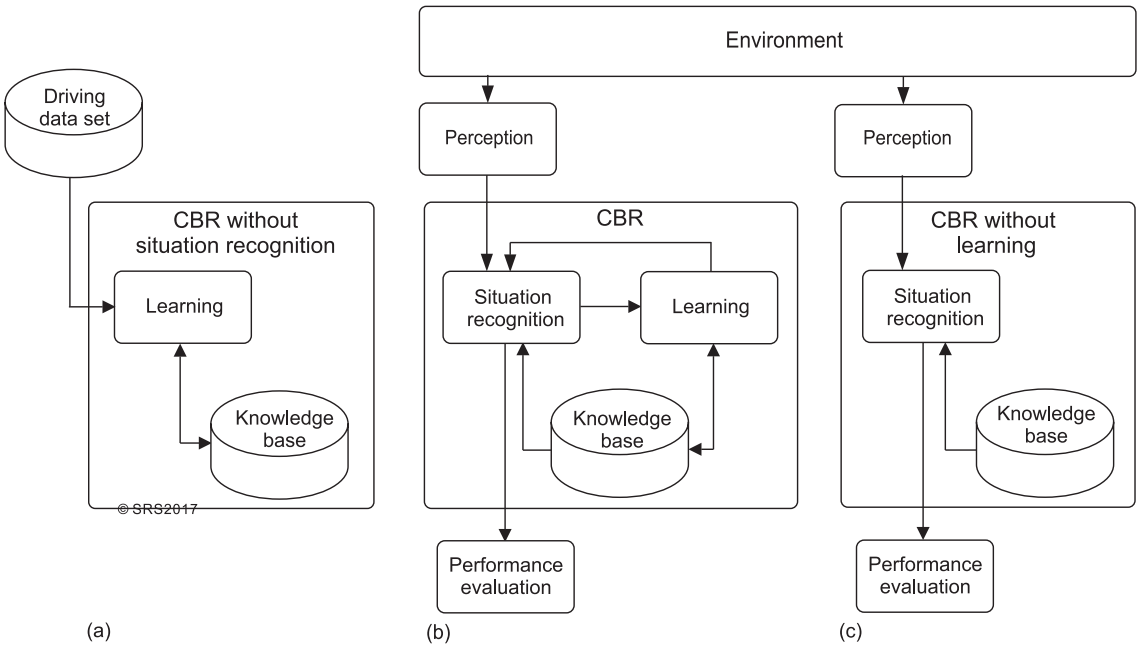

Figure 5.6: Different phases of experiments on driving data; (a) offline training phase, (b) online training phase, and (c) test phase

The case base and the database DB_{FDP} are initialized for each driving data through offline training phase using an entire training data set. In this phase, the proposed approach is evaluated using different test driving data related to 9 drivers individually. In each experiment, the knowledge base is initialized with a set of lane-change maneuvers performed by each driver. The occurring situations are labeled to set the situation patterns statically for training.

Relative changes in the values of two characteristics "heading angle between ego-vehicle and road" or "steering wheel angle" may represent lane-change maneuvers intended by human drivers.

The lane-change maneuvers are labeled by considering the changes in heading angle between ego-vehicle and road. The changes in steering wheel angle and the heading angle through a driving maneuver are investigated for different drivers. The investigations roughly present comparable results illustrated for one test driver in Fig. 5.7, when the driver is driving on the left or right lanes.

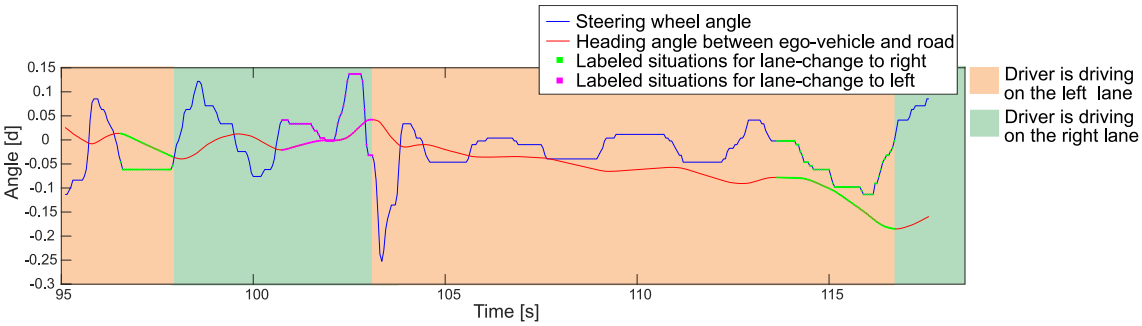

Figure 5.7: Changes in steering angle and heading angle through driving maneuvers

The blue line shows the oscillations of steering wheel angle changes along the driving maneuver. The swift changes in steering wheel angle are not reliable for determination of lane-change situations. Although steering wheel changes may represent a lane-change maneuver, this may even show a lane-keeping maneuver on a winding road.

In addition, the red line shows the changes in heading angle between ego-vehicle and road along the driving maneuver. Here, the colored zones where the driver is changing the driving lane, as well as the relation between slight changes in heading angle and occurring lane-change situations could be focused. A relationship between the changes in heading angle and occurring lane-change situations can be assumed. The situations between the last highest point or lowest point in the curve and the lane-change point could be marked as lane-change situations (shown for lane-change to right (green) and lane-change to left (pink)).

The labeling process could be executed to label the occurring events when an exact lane-change point (addressed as P_3 in Fig. 5.4) occurs. The labeling procedure configured for target application is presented in Algorithm 5. In this algorithm, τ_a shows heading angle variation threshold which is defined here by human experts.

Algorithm 5

```
procedure LABELING(List_sequence)
    //List_sequence : a sub sequence of occurring situations and operators
    LK ← 0;
    LCL ← 1;
    LCR ← 2;
    i ← 1;
    while i > size(List_sequence) do
        S_i ← List_sequence(i);
        S_{i+1} ← List_sequence(i + 1);
        S_i.label ← LK;
        if (S_i.c_2) < (S_{i+1}.c_2) then
            j ← i;
            while (S_j.c_25) <> LastLowestPoint((S_i.c_25), τ_a) do
                S_j ← List_sequence(j);
                S_j.label ← LCL;
                j ← (j − 1);
            end whileend;
            i ← (j + 1);
        end ifend;
        if (S_i.c_2) > (S_{i+1}.c_2) then
            j ← i;
            while (S_j.c_25) <> LastHighestPoint((S_j.c_25), τ_a) do
                S_j ← List_sequence(j);
                S_j.label ← LCR;
                j ← (j − 1);
            end whileend;
            i ← (j + 1);
        end ifend;
        i ← (i + 1);
    end whileend;
end procedureend.
```

The online training phase is executed through a complete fuzzy SOM-based CBR cycle detailed in sections 4.2 and 4.3 to adapt new patterns dynamically for the data became available in a sequential order. In this phase, situation recognition is evaluated at runtime, and the knowledge base is updated dynamically.

In the test phase, the accuracy of the proposed learned situation recognition approach is evaluated after online training. Driving duration considered for offline and online training as well as test with the number of lane-change maneuvers are given for each driver in Table 5.2.

Table 5.2: Experiments done by test drivers

Test drivers	Driving duration [minute:second]		
	Offline training	Online training	Test
#1	22:35	30:00	06:00
#2	21:43	24:00	05:00
#3	30:10	22:00	08:00
#4	26:12	35:00	05:00
#5	21:00	26:00	06:00
#6	26:23	24:00	07:00
#7	21:40	23:00	05:00
#8	28:02	28:00	04:00
#9	30:30	31:00	06:00

5.3.4 Clustering the driving behaviors

Based on the data set used for offline training and the values of the characteristic "time to collision" in three points P_1, P_2, and P_3, lane-change behaviors are divided into three clusters (A, B, and C). In this clustering, time to collision to "the vehicle in front" and "the vehicle right-behind" are considered for specification of lane-change to left and right behaviors alternatively. The clustering results of around 530 samples of lane-change to left and right maneuvers are graphically shown using Fig. 5.8 and Fig. 5.9. According to Fig. 5.8, cluster A shows the behaviors in which a driver intends to change the lane and finishes the maneuver when TTC is short (less than 9 seconds). A driver with the behavior cluster B starts and finishes a lane-change maneuver when the TTC is long. According to behavior cluster C, a driver intends to change the lane when it TTC is long and finished the maneuver when TTC is short. According to Fig. 5.9 addressing different behaviors of lane-change to the right, behavior cluster A is related to the drivers who start the maneuvers when TTC is more than 12, and change the lane when TTC is short. Cluster B indicates the behaviors in which the drivers mostly start and finish the maneuvers when TTC is low. In addition, cluster C including a few number of samples is related to the behaviors in which the TTC in the point P_2 is more than 12 seconds while TTC in other points is shorter.

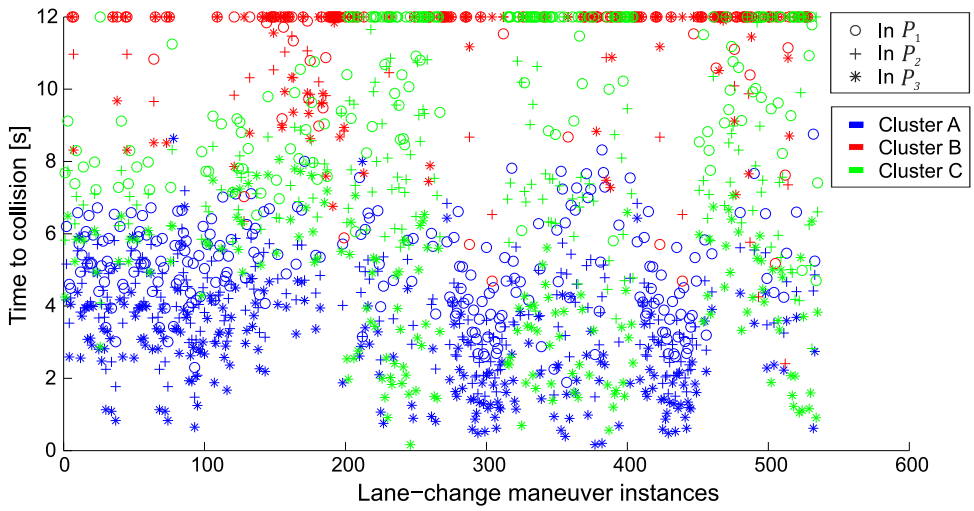

Figure 5.8: Different clusters of lane-change to left behaviors based on the drivers experiences

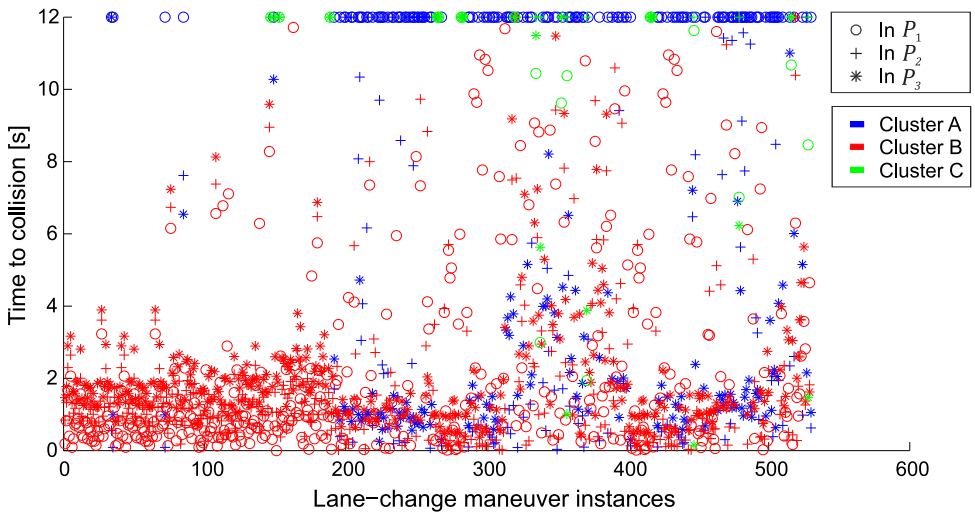

Figure 5.9: Different clusters of lane-change to right behaviors based on the drivers experiences

Here, it is supposed that the drivers with more than 50% experiences of one driving behavior cluster belong to that cluster. Accordingly, 5 different groups of drivers with a combination of defined lane-change clusters could be assigned as shown in Table 5.3.

Table 5.3: Percentage of experiments of test drivers for different behaviors

Test	Lane-change to left				Lane-change to right				Group
no.	A	B	C	Cluster	A	B	C	Cluster	name
#1	65.78	13.15	21.05	A	3.94	96.05	0	B	G1
#2	68.57	8.57	22.85	A	1.42	98.57	0	B	G1
#3	17.02	55.31	27.65	B	2.22	91.11	6.66	B	G2
#4	24.32	16.21	59.45	C	50.68	49.31	0	A,B	G3
#5	74.46	25.53	0	A	25.53	65.95	8.51	B	G1
#6	34.44	14.44	51.11	C	37.07	56.17	6.74	B	G4
#7	77.77	22.22	0	A	26.66	68.88	4.44	B	G1
#8	0	51.39	49.58	B,C	49.05	50.16	0	A,B	G5
#9	21.21	24.24	54.54	C	33.65	60.12	6.25	B	G4

5.3.5 Setting

For the implementation of fuzzy SOM-based CBR, different algorithms are applied to the CBR platform. The algorithm k-NN is applied for case retrieval. A selected feature selection algorithm and fuzzy density clustering method are applied to define relevant characteristics and modeling parameters of membership functions for those characteristics.

5.4 Evaluation results

5.4.1 Fuzzy SOM-based case representation

Description

Generation of membership functions for continuous variables plays an important role in case representation and individualization of driving behaviors. Generation of membership functions may be mostly done using human expert knowledge, which is neither practical nor reliable. As discussed in Section 4.4, different methods have been formerly proposed for the automatic generation of membership functions. According to [NU09], the density-based method with FN-DBSCAN shows high accuracy and low computational time. Additionally, based on the previous study [SHN-Son] on the application of the simulated driving data, FN-DBSCAN outperforms other methods in terms of membership functions generation time and accuracy. Accordingly, FN-DBSCAN is applied for generation of membership functions in this work. In addition, trapezoidal membership function which could outperform the triangular and Gaussian membership functions in pattern recognition [SBK16], is considered for defining the linguistic variables.

The applied FN-DBSCAN receives a stream of the gained values for each quantitative variable as input. It generates a set of trapezoidal membership functions for each variable and returns the cores and supports of the functions as output. Here, a crisp value of each quantitative variable is established through a set of its corresponding fuzzy values measured through the generated membership functions.

Results

According to Table 5.1, 21 quantitative characteristics could be represented with fuzzy values. to define a situation. These characteristics could be divided into 6 general types of variables according to their nature in using the same linguistic modifiers. Those variable types are: velocity of ego-vehicle Var_1, distance to other vehicles Var_2, velocity of other vehicles Var_3, time to collision Var_4, vehicle angle Var_5, and steering wheel angle Var_6.

By application of FN-DBSCAN, the number of generated membership functions for each variable is specified for the drivers (as shown in Table 5.4). According to Table 5.4, different amounts of membership functions are generated for the drivers with different behaviors. As an example, different behaviors of two drivers during lane-change maneuvers are presented in Fig. 5.10. According to the data density of TTC value shown in those figures, driver #7 tries mostly to change the lane when TTC is less than 5 seconds while driver #3 changes the lane cautiously when TTC is more than 4 seconds. These differences between the experiences affect the number and modeling parameters of membership functions for knowledge representation.

Table 5.4: Number of automatically generated membership functions for each variable type according to the gained values in drivers experienced maneuvers

Test drivers	Var_1	Var_2	Var_3	Var_4	Var_5	Var_6
#1	17	20	14	16	17	11
#2	18	14	14	15	12	11
#3	5	19	3	7	6	16
#4	15	19	3	8	6	16
#5	17	19	17	15	16	15
#6	19	18	5	8	6	6
#7	17	20	14	16	17	11
#8	8	14	13	13	12	5
#9	19	18	5	8	6	6

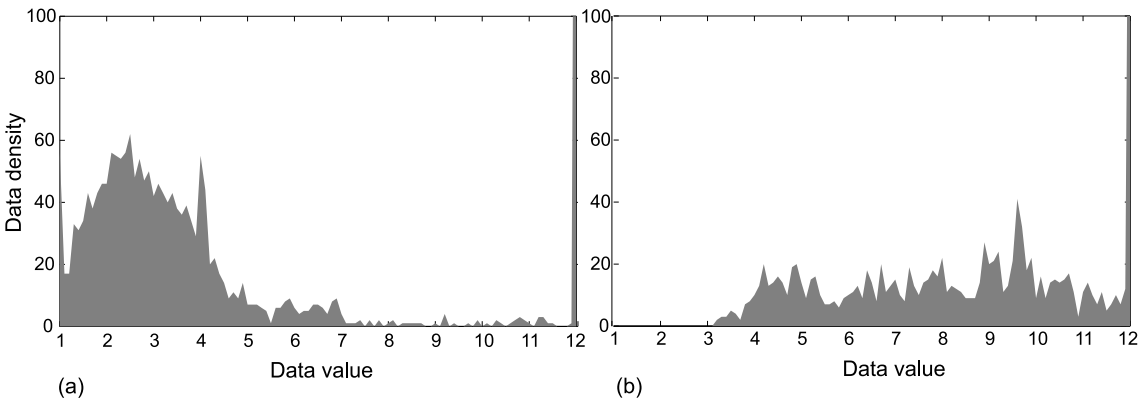

Figure 5.10: Density of TTC values experienced during the lane-change maneuvers performed by two drivers; (a) driver #7 and (b) driver #3

According to the defined parameters in DB_{FDP}, each quantitative characteristics of the occurring situations will be transformed to its corresponding fuzzy value. A sequence of occurring situations with the transformed fuzzy values is illustrated in Fig. 5.11. Here, steering wheel angle and accelerator pedal are considered as a combined operator. The value of each characteristic in Fig. 5.11 shows its corresponding linguistic value. Moreover, Eq. 5.9 addresses the generated cases from a sequence of a situation, trigger action, and the upcoming situation.

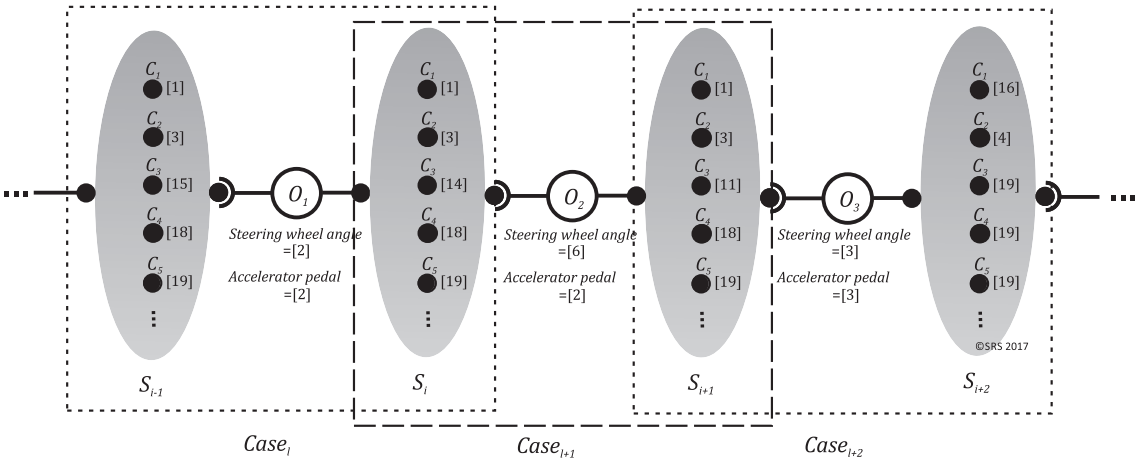

Figure 5.11: A sequence of occurring situations and the related fuzzy cases while generated

$$\begin{aligned} &Case_l \quad : if\ (S_{i-1})\ \&\ (O_1)\ then\ S_i;\ end; \\ &Case_{l+1} : if\ (S_i)\ \&\ (O_2)\ then\ S_{i+1};\ end; \\ &Case_{l+2} : if\ (S_{i+1})\ \&\ (O_3)\ then\ S_{i+2};\ end. \end{aligned} \tag{5.9}$$

5.4.2 Feature/characteristics weighting and selection

Description

The rough set theory is employed in this section to define a subset of relevant characteristics by a rank ordering of characteristics. In this section, the feature selection algorithms proposed in Section 4.5 are developed and evaluated. The evaluation is done in two steps. At first, the effectiveness of feature selection algorithms on situation recognition accuracy is measured. Here, each algorithm is applied to the learning process for feature selection. In the second step, the most suitable algorithm is selected and applied to the proposed framework for the next applications.

The conditional characteristics applied to the algorithm are given in Table 5.1. Three decision characteristics are defined based on the three situation patterns. Accordingly, a vector of impact factors W_s is generated for each human driver.

Results

A comparison between the proposed feature selection algorithms is shown in Table 5.5. This table gives the results about the percentage of the features selected using the proposed algorithm and the average situation recognition accuracy measured by applying those selected features. According to the results, the accuracy of the situation recognition without selection of relevant features is low. This shows the irrelevant features may decrease the recognition accuracy. Thus, selection of key characteristics for definition and identification of situation patterns as well as data reduction by ignoring the irrelevant characteristics is an important task.

Among four feature selection algorithms, the best performance in term of accuracy is achieved mostly by applying ST and HS strategies. However, the number of features in the selected sub-set by ST is smaller than this number by HS. Accordingly, the maximum accuracy and a small sub-set of the relevant features could be achieved by application of ST strategy.

In the next step, the rough set theory with application of ST strategy is considered in the proposed fuzzy SOM-based CBR. By applying the proposed feature selection for the data sets of the test drivers, the importance of each characteristic for the definition of lane-change situations is measured as shown in Fig. 5.12.

Table 5.5: Average recognition results using different selection algorithms

Driver	Feature selection (%)					Accuracy (%)				
set	None	QR	MS	HS	ST	None	QR	MS	HS	ST
#1	100	67.1	63.2	64.6	57.6	68.2	67.2	96.2	92.3	98.1
#2	100	55.7	65.3	74.2	65.2	81.4	89.4	65.1	95.0	96.5
#3	100	64.2	47.3	81.6	61.4	67.1	71.2	82.0	96.0	95.4
#4	100	56.4	52.9	86.9	72.9	82.7	77.4	94.2	93.2	94.9
#5	100	68.4	40.1	74.2	69.1	58.9	96.0	74.2	98.6	98.1
#6	100	40.1	76.4	92.1	80.6	78.5	76.2	95.0	94.5	93.2
#7	100	79.4	72.5	75.3	80.6	76.8	92.6	85.0	97.8	95.8
#8	100	61.4	32.9	74.9	65.2	85.2	96.7	65.1	95.9	97.5
#9	100	64.2	72.1	84.6	72.9	84.1	82.2	83.3	92.2	95.3

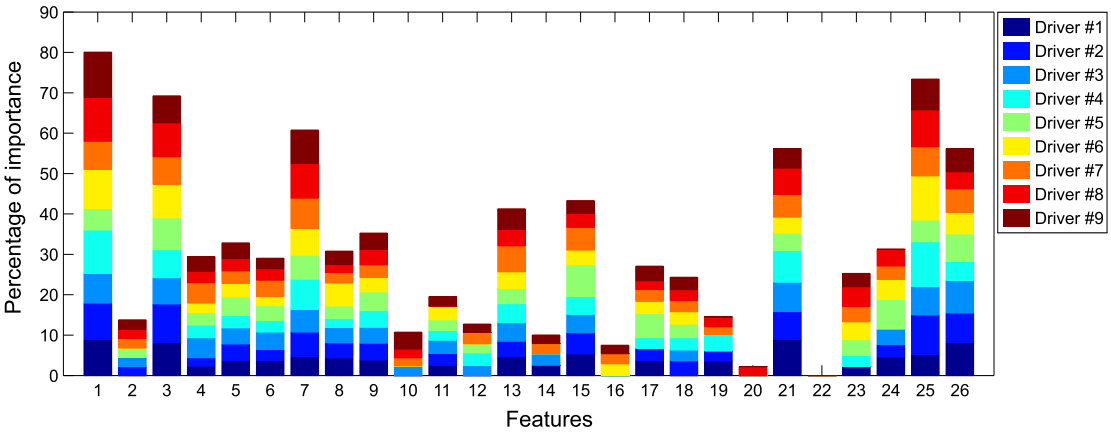

Figure 5.12: Average importance of each characteristic in recognition of the situations for different drivers

It could be detected in the figure that the measured W_s for different drivers are not equal. However, some of the characteristics with the impact factor of zero, do not contribute to the significance of the situation description and therefore can be neglected in indexing and similarity assessment.

5.4.3 10-Fold cross-validation

In this section, offline training process of the situation recognition approach is evaluated using the data informed in Table 5.2. The performance using 10-fold cross-validation could be measured by comparing the performance of each fold (see Fig. 5.13, Fig. 5.14, and Fig. 5.15). The shown results in Fig. 5.13 give the detection rates achieved for all 10 folds of each driver. In addition, Fig. 5.14 and Fig. 5.15 indicate the false alarm rate and accuracy resulting for all the folds for 9 drivers. The

results indicate that the proposed situation recognition approach is generalizable because the discrimination is not significant.

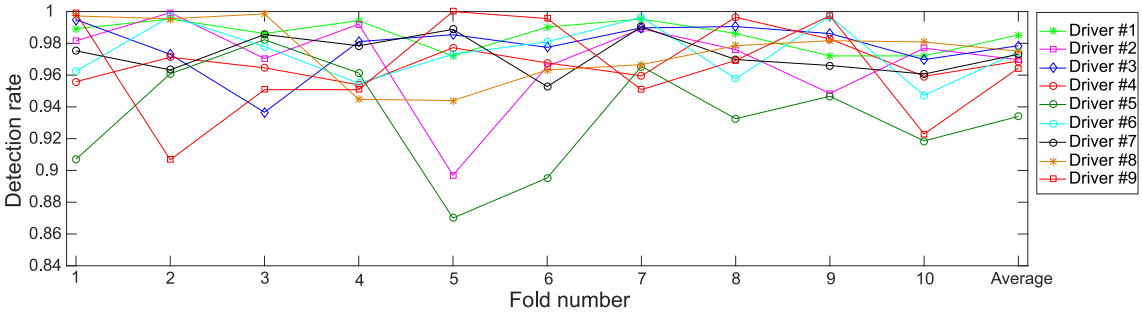

Figure 5.13: Average detection rate measured through 10-fold cross-validation for different drivers

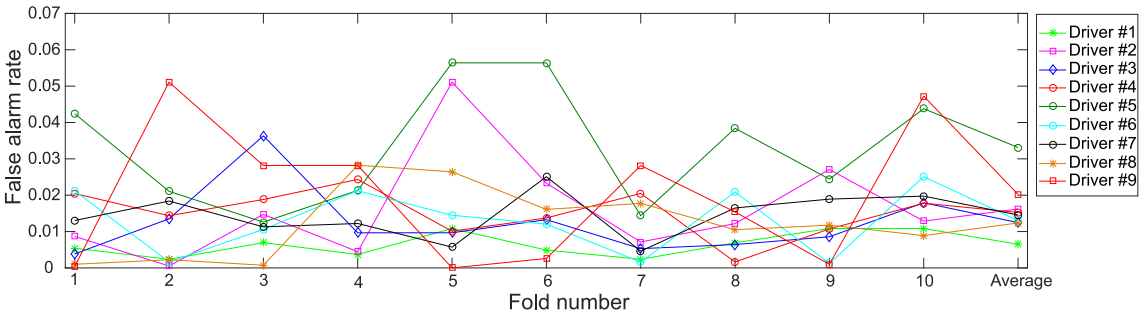

Figure 5.14: Average false alarm rate measured through 10-fold cross-validation for different drivers

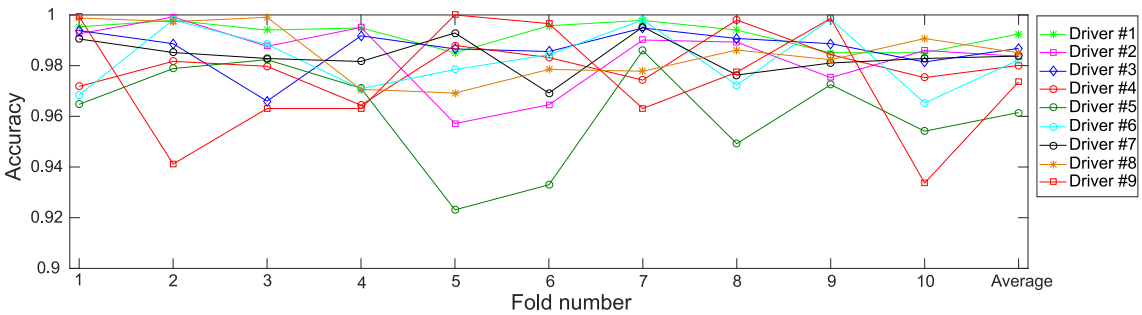

Figure 5.15: Average accuracy measured through 10-fold cross-validation for different drivers

Accordingly, it can be detected from Fig. 5.13 that the average accuracy of 10-fold cross-validation is more than 93%. In addition, detection rate and false alarm rate point an acceptable situation recognition results.

5.4.4 Learning performance

Description

Here, the developed approach is considered to assess online learning performance. For each driver, the driver specified approach is used and connected with the related individual knowledge base. The knowledge base would be updated for each new lane-change experience when an occurring situation is not similar to the existing experiences/cases and recognized incorrectly. Here, situation recognition performance is evaluated in terms of detection rate, false alarm rate, and recognition accuracy.

Results

Evaluation results of the online training process for all test drivers are illustrated in Fig. 5.16. The plot is representing the accuracy of situation recognition in around 20 minutes driving through the online training phase. The situation recognition results in the first minutes of driving confirm the importance of the online learning. Considerable changes in the measured situation recognition accuracy can be detected.

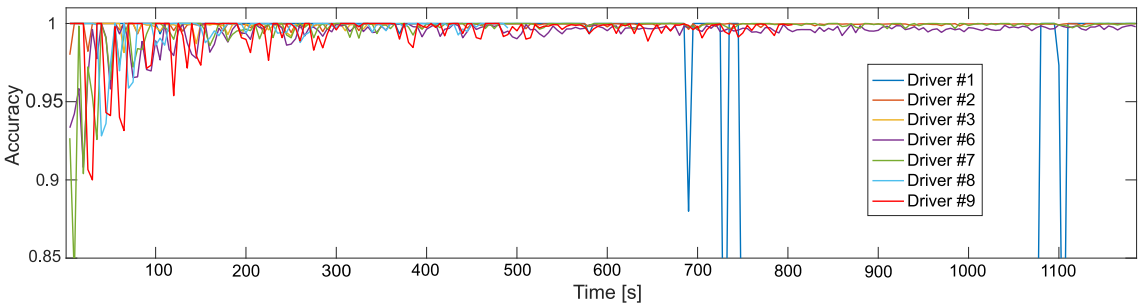

Figure 5.16: Situation recognition performance obtained (every 5 seconds) during online training for around 20 minutes driving by all the drivers

In Fig. 5.16, a significant difference among the accuracy curve behaviors is shown. As it could be detected for most of the drivers, the situation recognition performance significantly increases in the next minutes. However, the performance is not stable and may slightly change. Another behavior could be detected from driver #1 in

this experiment. The performance is changed considerably several times in the last minutes. The average elapsed time for learning is measured 0.9 seconds in average.

Now, those two behaviors would be considered more in detail by measuring the percentage of knowledge base workload during online learning. Knowledge base workload is measured by determining the number of queries executed on the case base or required for updating the knowledge base in a given period of time.

Those behaviors are presented graphically in Fig. 5.17 to compare the performance changes during online training and knowledge base workload. By comparing Fig. 5.17 (a) and (b) related to two different drivers, it could be seen that the performance and workload change nearly at the same time. However, low performance of situation recognition in the last minutes of driving by driver #1 could be detected.

By comparing the time of performance and workload changes (for example as focused in Fig. 5.17 (c) and (d)), it could be seen that workload changes follow the changes in situation recognition accuracy. It shows that decreasing the performance because of occurring new situations causes a set of transactions to update the related databases. Accordingly, low performance of situation recognition in the last minutes and high knowledge base workload related to driver #1 are because of a large number of transactions needed for updating the case base and databases for new experiences. The obtained results emphasize the significance of online learning.

In this step, the developed fuzzy SOM-based CBR is trained through offline training and evaluated during and after online training for each driver. Hereby, the driving data noted in Table 5.2 is considered for learning and test. Detection rate and false alarm rate measured for situation recognition are addressed in Table 5.6.

According to the results addressed in Table 5.6, considerable improvements in recognition performance after learning process could be seen. Detection rate for most of the cases is improved. However, false alarm rate in all cases is reduced to less than 0.02.

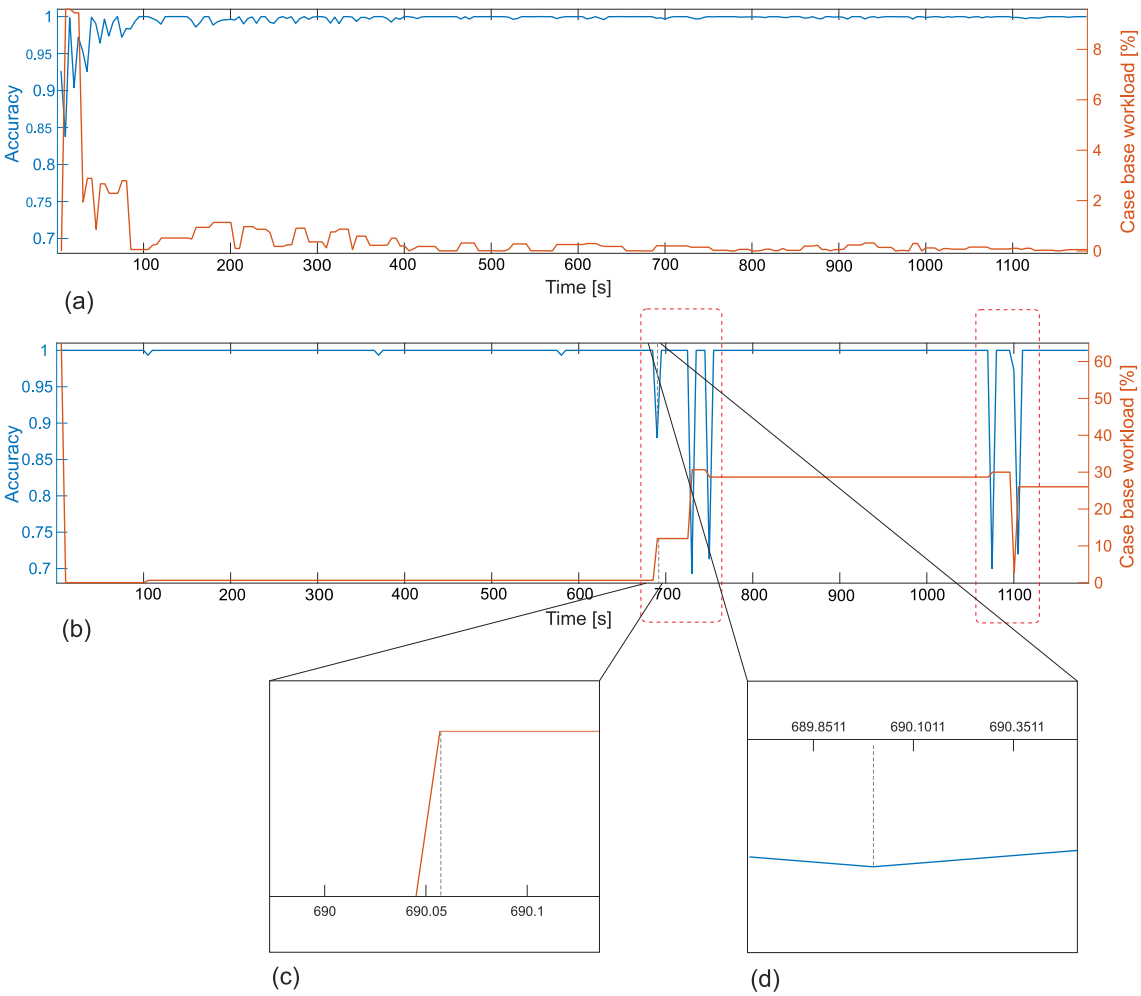

Figure 5.17: Situation recognition performance obtained (every 5 seconds) during online training for around 20 minutes driving (a) by driver #7 (b) by driver #1

5.4.5 Situation recognition performance

Description

In this section, the performance of the proposed approach is discussed. The performance is detailed for three situation patterns introduced in Section 5.3.2.

In addition, this is an attempt to answer the question, "Can the performance of situation recognition be improved by considering individual behaviors of drivers?". Here, individual behaviors are defined by the knowledge represented individually in different knowledge bases. To approach this question, the performance of situation recognition is measured for the test driver data in two conditions

Table 5.6: Evaluation of situation recognition performance in average before and after online training for each driver

	Before online training		After online training	
Driver	DR	FAR	DR	FAR
#1	0.9522	0.0760	**0,9794**	**0,0143**
#2	0.9678	0.0465	**0.9756**	**0.0129**
#3	0.9714	0.0371	0.9646	**0.0179**
#4	0.9683	0.0335	**0.9886**	**0.0074**
#5	0.9551	0.0563	**0.9654**	**0.0175**
#6	0.9755	0.0335	0.9671	**0.0147**
#7	0.9872	0.0212	0.9770	**0.0097**
#8	0.9729	0.0352	**0.9883**	**0.0048**
#9	0.9585	0.1187	0.9550	**0.0217**

(a) the test driver data evaluation is based on their personalized knowledge base and

(b) the test driver data evaluation is based on the knowledge base trained by other drivers.

Results

To address the system output in the recognition process, a test driving sequence with the duration of around 4 minutes for one driver is shown in Fig. 5.18. The changes in the some characteristics values in occurring situations are given. To highlight the changes in characteristics values clearly, some characteristics are chosen to be shown in this figure. In addition, relative velocities are illustrated instead of the velocity of the vehicles. The situation S_{i} in Fig. 5.18 is defined as a set of characteristics C with their actual values modeled using SOM. Through the recognition process, the lane-change situations which are recognized with an average response time 0.04 seconds are marked in green. Also, the real lane changing maneuvers realized by the human driver are marked in red. The result indicates detection rate of the proposed approach and shows that the lane changes realized by the human driver are also detected by the system. Additionally, in each detected lane changing situation, the similar experienced situations and the percentage of their usage by retrieve process are partly addressed in Fig. 5.18. This percentage shows the applicability of the cases in various situations (coverage rate). A case with higher coverage rate is more generalized and applicable to a wide range of situations.

Moreover, situation recognition accuracy measured for the driving data is shown in Table 5.7. The accuracy is calculated for different situation patterns by considering the two defined conditions (a) and (b). Here, the human drivers who are using

situation recognition framework without learning process are called KB users. In addition, the drivers who are using the developed framework with learning ability and therefore the knowledge base is adapted using their experiences are called KB trainer. According to the average accuracy, the situation recognition in the case when human drivers are using their knowledge base is acceptable (more than 0.92). In most of the experiments, this value is higher than the measured situation recognition in the case when drivers are not using their experienced knowledge base. However, recognition accuracy decreases when human drivers do not apply their individualized knowledge base.

Table 5.7: Evaluation of situation recognition performance using fuzzy SOM-based CBR

Driver as KB trainer	Situation pattern	Drivers as KB users								
		#1	#2	#3	#4	#5	#6	#7	#8	#9
#1	LK	0.9913	0.9723	0.9098	0.6825	0.9452	0.6614	0.8934	0.6466	0.7467
	LCL	0.9896	0.9868	0.9536	0.8039	0.9763	0.8316	0.9004	0.8016	0.8919
	LCR	0.9941	0.9856	0.9561	0.8394	0.9672	0.7885	0.9312	0.8312	0.8398
	Ave.	**0.9917**	**0.9815**	0.9398	0.7753	0.9629	0.7605	0.9083	0.7598	0.8261
#2	LK	0.9510	0.9629	0.9211	0.6850	0.9557	0.6294	0.9167	0.6420	0.7239
	LCL	0.9754	0.9781	0.9526	0.7993	0.9775	0.8211	0.9126	0.7947	0.9040
	LCR	0.9741	0.9848	0.9685	0.8445	0.9767	0.7701	0.9391	0.8346	0.7964
	Ave.	0.9668	**0.9753**	0.9474	0.7763	**0.9700**	0.7402	0.9228	0.7571	0.8081
#3	LK	0.8688	0.9102	0.9765	0.7107	0.7982	0.6886	0.8295	0.6772	0.7317
	LCL	0.9076	0.9397	0.9822	0.8296	0.8511	0.8587	0.8842	0.8245	0.8421
	LCR	0.9606	0.9705	0.9941	0.8380	0.9471	0.7923	0.9451	0.8228	0.8633
	Ave.	0.9123	0.9402	**0.9842**	0.7928	0.8655	0.7798	0.8863	0.7748	0.8124
#4	LK	0.6359	0.6884	0.7168	0.9054	0.7313	0.5672	0.7472	0.7185	0.5655
	LCL	0.8647	0.8772	0.9485	0.9162	0.7807	0.8099	0.7634	0.8020	0.8593
	LCR	0.7507	0.7968	0.7683	0.9849	0.9122	0.6938	0.8758	0.8835	0.6722
	Ave.	0.7504	0.7874	0.8112	**0.9355**	0.8080	0.6903	0.7955	0.8013	0.6990
#5	LK	0.8835	0.9009	0.8636	0.6951	0.9813	0.5514	0.9140	0.6550	0.6272
	LCL	0.9600	0.9610	0.9550	0.7918	0.9890	0.8072	0.9188	0.7911	0.8916
	LCR	0.9211	0.9399	0.9086	0.8696	0.9923	0.6944	0.9507	0.8546	0.7149
	Ave.	0.9216	0.9339	0.9091	0.7855	**0.9875**	0.6843	0.9278	0.7669	0.7446
#6	LK	0.6335	0.6183	0.7204	0.6071	0.6549	0.9076	0.6176	0.5638	0.9345
	LCL	0.6867	0.6527	0.7605	0.7161	0.6772	0.9463	0.7103	0.7254	0.9756
	LCR	0.9405	0.9633	0.9599	0.8443	0.9778	0.9268	0.8156	0.8032	0.9441
	Ave.	0.7536	0.7448	0.8136	0.7225	0.7699	**0.9269**	0.7145	0.6975	**0.9514**
#7	LK	0.9510	0.9170	0.5879	0.7722	0.9673	0.5163	0.9616	0.6935	0.4377
	LCL	0.9780	0.9730	0.9488	0.8189	0.9880	0.8142	0.9563	0.8127	0.8838
	LCR	0.9678	0.9436	0.6312	0.8943	0.9793	0.6049	0.9553	0.8505	0.4983
	Ave.	**0.9656**	0.9445	0.7226	0.8285	**0.9782**	0.6451	**0.9577**	0.7856	0.6066
#8	LK	0.5837	0.5916	0.3877	0.7129	0.6939	0.5689	0.6297	0.9784	0.4546
	LCL	0.8230	0.7920	0.6607	0.7568	0.7664	0.7797	0.6318	0.9854	0.6473
	LCR	0.7560	0.7819	0.7067	0.8923	0.9275	0.6798	0.8724	0.9902	0.7076
	Ave.	0.7209	0.7218	0.5850	0.7873	0.7959	0.6761	0.7113	**0.9847**	0.6032
#9	LK	0.7550	0.7908	0.8061	0.6732	0.7569	0.9530	0.6838	0.6431	0.9311
	LCL	0.8793	0.8902	0.9207	0.7715	0.7789	0.9669	0.7503	0.7870	0.9674
	LCR	0.8698	0.8970	0.8854	0.8371	0.9780	0.9696	0.8192	0.8479	0.9552
	Ave.	0.8347	0.8593	0.8707	0.7606	0.8379	**0.9632**	0.7511	0.7594	**0.9512**

The performance of individualization is further evaluated by comparing the situation recognition in two conditions: (c) each driver uses an individualized knowledge base initialized by the experiences of the same group drivers, (d) each driver uses a general knowledge base initialized using the whole experiences of all drivers. The evaluation results in terms of detection rate and false alarm rate in both conditions for recognition of three situation patterns are shown in Fig. 5.19.

As illustrated in Fig. 5.19, the detection rate for lane-keeping is improved for the drivers when an individualized situation recognition is applied. Lane-change detection rate is improved for most of the drivers in condition (c) except drivers #1 and #2. It is maybe because of using the experiences of more drivers in the same group with drivers #1 and #2. Further evaluation of recognition results indicates that false alarm rate mostly decreases by applying an individualized knowledge base.

As shown in Fig. 5.19, false alarm rate decreases for all the test drivers in lane-change to right situation recognition by applying a related individual knowledge base. The same results could be detected mostly for lane-keeping and lane-change to right. However, false alarm rate evaluation results highlight the effectiveness of using individualized situation recognition compared to a generalized situation recognition.

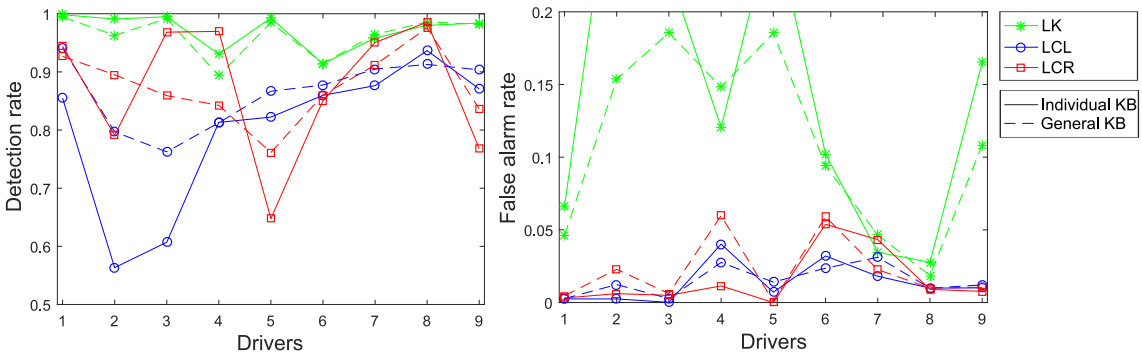

Figure 5.19: Evaluation of the proposed approach by comparing individualized and general knowledge base for situation recognition

Another attempt to express the effectiveness of individualized lane-change situation recognition is considered here. The proposed framework is trained (through online learning) by a driver of Group A. Then, the framework is used by two test drivers of Groups A and B separately. This is an attempt to answer the question, "Can the performance of situation recognition be improved by consideration of individual behaviors of drivers?".

Results of the recognized lane-change situations in the first and second tests are shown in Fig. 5.20 and Fig. 5.21. The blue lines show the driving lane of ego-vehicle

at time. The points marked in green and blue indicate the labeled lane-change situations and recognized lane-change situations alternatively. The red circles in the figures could obviously highlight the visible situations which are recognized incorrectly as well as the situations which are not detected correctly. The results presented in Fig. 5.20 state that the recognized situations nearly overlap the labeled situations. However, the recognized situations illustrated in Fig. 5.21 confirm the inconsistency between the recognized and labeled situations. Using an unrelated knowledge could increase false alarm rate.

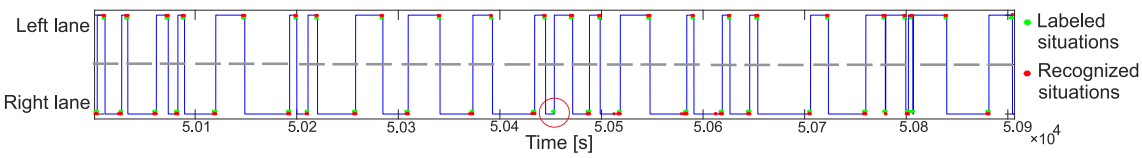

Figure 5.20: The lane-change situations recognized for a driver of group A using the assistance system trained for the same group of drivers

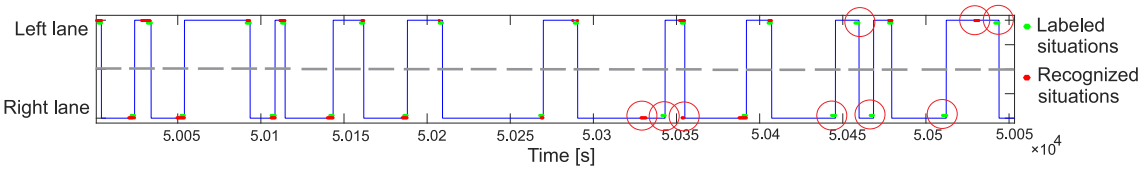

Figure 5.21: The lane-change situations recognized for a driver of group B using the assistance system trained for drivers of group A

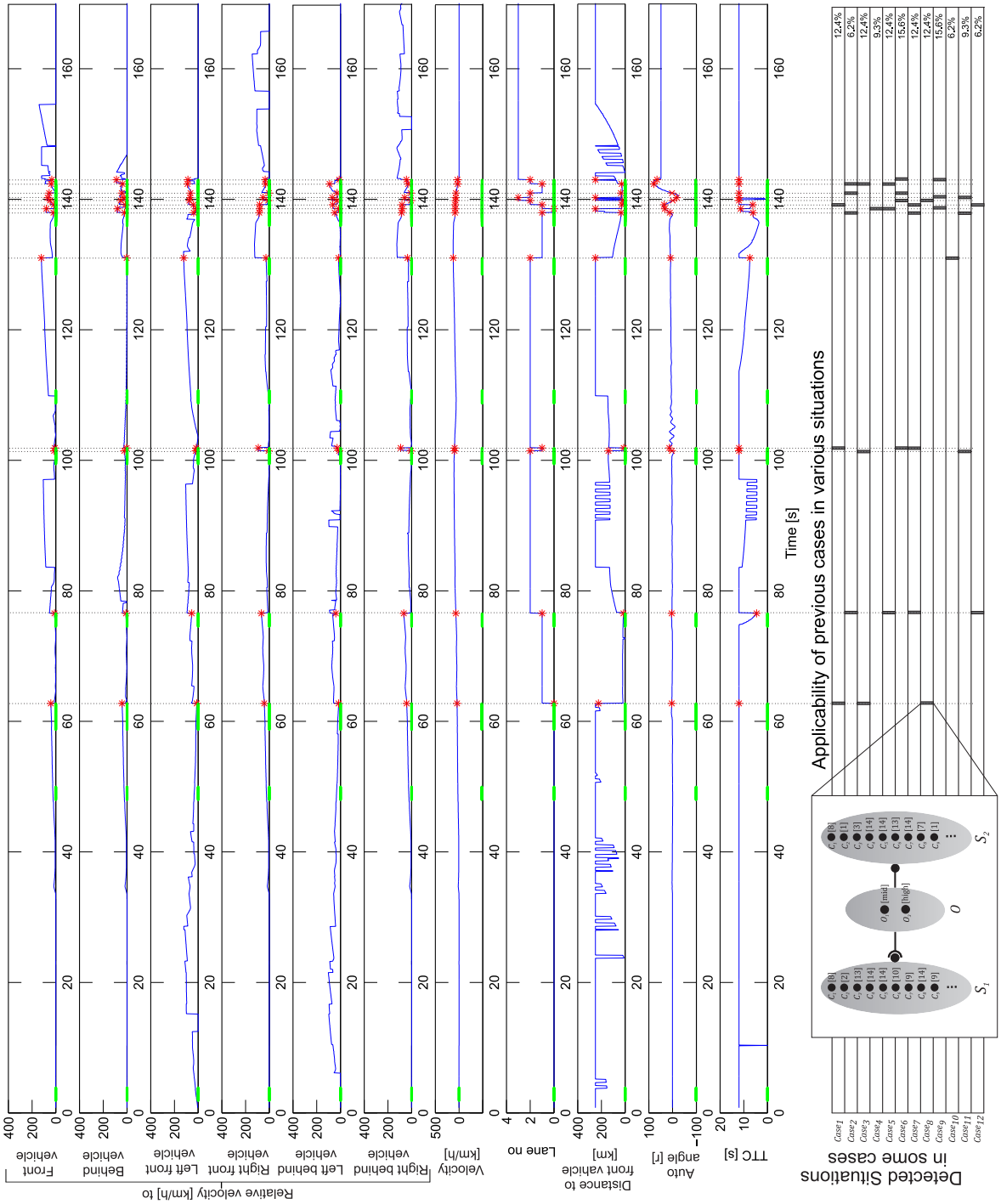

Figure 5.18: Recognized lane-change situations by the framework related to a test driver with a driving duration of 4 minutes (the recognized and real experimented lane-change situations are indicated in green and red respectively)

6 Summary, conclusion, and future work

6.1 Summary and conclusion

In this work, a new framework is proposed and developed for individualized situation recognition. The main idea is developing an approximate reasoning approach driven by learning and reusing human-operator experiences to handle event-discrete situations in unknown dynamic environments.

An improved CBR approach is applied for the realization of individualized situation recognition. A knowledge representation approach by integration of SOM with FL is proposed for the first time to support the situation recognition process.

Through a survey presented in Chapter 2, the current status of the proposed approximate reasoning, fuzzy-based CBR, is reviewed. The advantages and disadvantages of the existing fuzzy-based CBR approaches are considered for generation of the new framework. The proposed fuzzy SOM-based CBR is addressed as an individualized approach by applying personalized knowledge of exclusive definitions of experienced events.

The proposed approach is developed and evaluated for supervision of human drivers to show the applicability and performance of the introduced approach. In the proposed approach, SOM is applied for modeling dynamical changes in the lane-change maneuvers as situations and operators. The main functionality of this supervision is to recognize lane-change driving situations while they are taking place without consideration of human intention. The proposed approach was used by 9 test drivers individually for initialization, offline training, online training, and test phases. In addition, different driving behavior groups for lane-change to left and right are defined to evaluate the effectiveness of individualization in situation recognition.

According to the validation and evaluation results, the research objectives related to the research questions designed in Section 1.2.1 could be achieved as follows

1. The experimental results show that the proposed situation recognition framework based on approximate CBR features a new performance quality. Event-discrete situations could be modeled and represented in the knowledge base. Results of cross-validation show generalizability of the proposed approach by applying unknown data for training and test.

2. A further evaluation of the performance of online learning and knowledge base adaption shows that situation recognition performance could be improved significantly with respect to the false alarm rate.

3. The situation recognition accuracy has been evaluated additionally for different groups of trainers and users of the knowledge base in fuzzy SOM-based CBR. The results indicate that the performance of situation recognition is enhanced when the knowledge base is personalized. Moreover, the performance of situation recognition by applying an individualized situation recognition could be improved significantly compared to a general situation recognition trained by the experiences of all drivers.

6.2 Research contributions

The main contributions identified as an outcome of this thesis are as follows

- a conceptualization and definition of situation recognition problem,
- individualized real-time situation recognition based on fuzzy SOM-based CBR,
- a new knowledge representation approach for modeling and individual characterization event-discrete situations,
- a new reasoning process to adapt the knowledge base by learning from new event-discrete situations.

Furthermore, some minor contributions can be identified in this work. Here, a fuzzy object-oriented knowledge base has been designed and structured to support the individualized situation recognition framework. A new approach by integration of SOM with FL has been proposed for repressing the events in dynamic environments. Different procedures are integrated with the designed framework for further improvements of situation recognition. An approach has been proposed for automatic generation of membership functions to transfer the value of continues variables from crisp to fuzzy. Moreover, an approach for feature selection has been presented to be used for case indexing and retrieval.

6.3 Future work

The proposed approach could be examined for different complex real-world applications. In addition, other characteristics affecting behaviors of human drivers in different situations such as the characteristics showing humanś emotional state could be considered in the definition of situations. Moreover, learning process could be improved by application of retention/memory and forgetting techniques.

The proposed situation recognition framework could also be integrated with an intention recognition process. Considering the intention of human operators as well as other participants in the environment may improve the supervision performance by decreasing the false alarm rate.

Bibliography

[AA12] A. Atanassov and L. Antonov. Comparative analysis of case based reasoning software frameworks jCOLIBRI and myCBR. *Journal of the University of Chemical Technology & Metallurgy*, 47(1):83–90, 2012.

[ACN02] S. Agrawal, S. Chaudhuri, and V. R. Narasayya. Index and materialized view selection for a given workload, April 2 2002. US Patent 6,366,903.

[AE13] J. R. Anaraki and M. Eftekhari. Rough set based feature selection: a review. In *5th Conference on Information and Knowledge Technology (IKT)*, pages 301–306. IEEE, 2013.

[AEGT04] E. Armengol, F. Esteva, L. Godo, and V. Torra. On learning similarity relations in fuzzy case-based reasoning. In *Transactions on Rough Sets II*, pages 14–32. Springer, 2004.

[AHY11] S. An, Q. Hu, and D. Yu. Case-based classifiers with fuzzy rough sets. *Rough Sets and Knowledge Technology*, pages 172–177, 2011.

[AIA10] A. Aztiria, A. Izaguirre, and J. C. Augusto. Learning patterns in ambient intelligence environments: a survey. *Artificial Intelligence Review*, 34(1):35–51, 2010.

[AIA17] S. A. AIAI. AIAI CBR Shell. Available at http://www.aiai.ed.ac.uk/project/cbr/CBRDistrib/ (last accessed: 2017/06/28), 2017.

[AL13] K. Amailef and J. Lu. Ontology-supported case-based reasoning approach for intelligent m-government emergency response services. *Decision Support Systems*, 55(1):79–97, 2013.

[AS14] N. Adhithyaa and P. Sumathi. A survey on retrieval strategy for case-based reasoning using modified-USIMSCAR for objected oriented cases. *International Journal of Advanced Research in Computer Science*, 5(7):150–154, 2014.

[AWS14] A. Amini, T. Y. Wah, and H. Saboohi. On density-based data streams clustering algorithms: a survey. *Journal of Computer Science and Technology*, 29(1):116–141, 2014.

[Baa03] F. Baader. *The description logic handbook: Theory, implementation and applications*. Cambridge university press, 2003.

[BAF+09] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and B. Von Schéele. A case-based decision support system for individual stress diagnosis using fuzzy similarity matching. *Computational Intelligence*, 25(3):180–195, 2009.

[BAF+11] S. Begum, M. U. Ahmed, P. Funk, N. Xiong, and M. Folke. Case-based reasoning systems in the health sciences: a survey of recent trends and developments. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(4):421–434, 2011.

[Bal79] J. F. Baldwin. A new approach to approximate reasoning using a fuzzy logic. *Fuzzy sets and systems*, 2(4):309–325, 1979.

[BBH+10] C. Bettini, O. Brdiczka, K. Henricksen, J. Indulska, D. Nicklas, A. Ranganathan, and D. Riboni. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2):161–180, 2010.

[BDPB13] S. K. Biswas, A. K. Das, B. Purkayastha, and D. Barman. Techniques for efficient case retrieval and rainfall prediction using cbr and fuzzy logic. *International Journal of Electronics Communication and Computer Engineering*, 4(3):692–698, 2013.

[Ber11] P. Berka. NEST: a compositional approach to rule-based and case-based reasoning. *Advances in Artificial Intelligence*, 2011:1–16, 2011.

[Bic08] I. Bichindaritz. Memory structures and organization in case-based reasoning. In *Case-Based Reasoning on Images and Signals*, pages 175–194. Springer, 2008.

[BKM+03] K. Baclawski, M. K. Kokar, C. J. Matheus, J. Letkowski, and M. Malczewski. Formalization of situation awareness. In *Practical Foundations of Business System Specifications*, pages 25–39. Springer, 2003.

[BKP05] R. Bergmann, J. Kolodner, and E. Plaza. Representation in case-based reasoning. *The Knowledge Engineering Review*, 20(03):209–213, 2005.

[BLNZ15] N. Bansal, K.-W. Lee, V. Nagarajan, and M. Zafer. Minimum congestion mapping in a cloud. *SIAM Journal on Computing*, 44(3):819–843, 2015.

[BMS16] A. Bartels, M.-M. Meinecke, and S. Steinmeyer. Lane change assistance. *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*, pages 1235–1257, 2016.

[BRS15] D. Borrajo, A. Roubíčková, and I. Serina. Progress in case-based planning. *ACM Computing Surveys (CSUR)*, 47(2):1–39, 2015.

[BS11] F. Bobillo and U. Straccia. Fuzzy ontology representation using OWL 2. *International Journal of Approximate Reasoning*, 52(7):1073–1094, 2011.

[BTGCDA04] J. J. Bello-Tomás, P. A. González-Calero, and B. Díaz-Agudo. Jcolibri: An object-oriented framework for building CBR systems. *Advances in case-based reasoning*, pages 29–39, 2004.

[BTRC+16] T. Bouabana-Tebibel, S. H. Rubin, A. Chebba, S. Bediar, and S. Iskounen. Knowledge induction based on randomization in case-based reasoning. In *IEEE 17th International Conference on Information Reuse and Integration (IRI)*, pages 541–548. IEEE, 2016.

[CAF12] P. Cingolani and J. Alcala-Fdez. jfuzzylogic: a robust and flexible fuzzy-logic inference system language implementation. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2012.

[CATU10] S. Chumklin, S. Auephanwiriyakul, and N. Theera-Umpon. Microcalcification detection in mammograms using interval type-2 fuzzy logic system with automatic membership function generation. In *IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 1–7. IEEE, 2010.

[CFD10] P.-C. Chang, C.-Y. Fan, and W.-Y. Dzan. A CBR-based fuzzy decision tree approach for database classification. *Expert Systems with Applications*, 37(1):214–225, 2010.

[CG11] P. Cadena and L. Garrido. Fuzzy case-based reasoning for managing strategic and tactical reasoning in starcraft. In *Mexican International Conference on Artificial Intelligence*, pages 113–124. Springer, 2011.

[CLBV08] S. Chandana, H. Leung, E. Bosse, and P. Valin. Fuzzy cognitive map based situation assessment for coastal surveillance. In *11th International Conference on Information Fusion*, pages 1–6. IEEE, 2008.

[CMS+10] O. Cominetti, A. Matzavinos, S. Samarasinghe, D. Kulasiri, S. Liu, P. Maini, and R. Erban. DifFUZZY: a fuzzy clustering algorithm for complex datasets. *International Journal of Computational Intelligence in Bioinformatics and Systems Biology*, 1(4):402–417, 2010.

[CMS14] Z. Cheng, H. Min, and X. Shen. Case-based reasoning with improved vector space model of indoor mobile robot. In *11th World Congress on Intelligent Control and Automation (WCICA)*, pages 2239–2246. IEEE, 2014.

[CNSZ11] J. L. Castro, M. Navarro, José M. S., and J. M. Zurita. Introducing attribute risk for retrieval in case-based reasoning. *Knowledge-Based Systems*, 24(2):257–268, 2011.

[CSW01] G. Cao, S. Shiu, and X. Wang. A fuzzy-rough approach for case base maintenance. In *International Conference on Case-Based Reasoning*, pages 118–130. Springer, 2001.

[CX04] S.-Y. Chiao and C. S. Xydeas. Using hierarchical HMMs in dynamic behavior modeling. *Proc. FUSION 2004*, pages 576–582, 2004.

[DAGCRGSRG07] B. Díaz-Agudo, P. A. González-Calero, J. A. Recio-García, and A. A. Sánchez-Ruiz-Granados. Building CBR systems with jCOLIBRI. *Science of Computer Programming*, 69(1):68–75, 2007.

[Dah11] A. Dahlbom. *Petri nets for situation recognition.* PhD thesis, Örebro university, Schweden, 2011.

[Dar09] A. Darwiche. *Modeling and reasoning with Bayesian networks.* Cambridge University Press, 2009.

[dBBL17] L. C. de Barros, R. C. Bassanezi, and W. A. Lodwick. Notions of fuzzy logic. In *A First Course in Fuzzy Logic, Fuzzy Dynamical Systems, and Biomathematics*, pages 53–78. Springer, 2017.

[DCGH12] J. Derrac, C. Cornelis, S. García, and F. Herrera. Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection. *Information Sciences*, 186(1):73–92, 2012.

[Dev06] K. Devlin. Situation theory and situation semantics. *Handbook of the History of Logic, Elsevier*, 7:601–664, 2006.

[DFE08] A. Davoust, M. W. Floyd, and B. Esfandiari. Use of fuzzy histograms to model the spatial distribution of objects in case-based reasoning. In *Conference of the Canadian Society for Computational Studies of Intelligence*, pages 72–83. Springer, 2008.

[DGG93] C. Dousson, P. Gaborit, and M. Ghallab. Situation recognition: representation and algorithms. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 166–172, 1993.

[DHO08] I. Derbel, N. Hachani, and H. Ounelli. Membership functions generation based on density function. In *CIS'08. International Conference on Computational Intelligence and Security*, pages 96–101. IEEE, 2008.

[DHP02] D. Dubois, E. Hullermeier, and H. Prade. Fuzzy set-based methods in instance-based reasoning. *IEEE Transactions on Fuzzy Systems*, 10(3):322–332, 2002.

[DPDRJ11] N. Douali, E. I. Papageorgiou, J. De Roo, and M.-C. Jaulent. Case based fuzzy cognitive maps (CBFCM): new method for medical reasoning: comparison study between CBFCM/FCM. In *IEEE International Conference on Fuzzy Systems (FUZZ)*, pages 844–850. IEEE, 2011.

[DSS93] R. Davis, H. Shrobe, and P. Szolovits. What is a knowledge representation? *AI magazine*, 14(1):1–17, 1993.

[EG00] M. R. Endsley and D. J. Garland. *Situation awareness analysis and measurement*. CRC Press, 2000.

[EM14] P. ElKafrawy and R. A. Mohamed. Comparative study of case based reasoning software. *Scientific Research and Management Studies*, pages 224–233, 2014.

[end] Situation awareness global assessment technique (sagat.

[End95] M. R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1):32–64, 1995.

[End16] M. R. Endsley. *Designing for situation awareness: An approach to user-centered design*. CRC press, 2016.

[fAI17] German Research Center for Artificial Intelligence. myCBR. Available at http://www.mycbr-project.net/ (last accessed: 2017/06/28), 2017.

[FB12] Y. Fischer and J. Beyerer. Defining dynamic bayesian networks for probabilistic situation assessment. In *15th International Conference on Information Fusion (FUSION)*, pages 888–895. IEEE, 2012.

[FDE08] M. Floyd, A. Davoust, and B. Esfandiari. Considerations for real-time spatially-aware case-based reasoning: A case study in robotic soccer imitation. *Advances in Case-Based Reasoning*, pages 195–209, 2008.

[Fis14] Y. Fischer. On situation modeling and recognition. In *Proceedings of the of the 2009 Joint Workshop of Fraun hofer IOSB and Institute for Anthromatics, Vision and Fusion Laboratory*, pages 203–215, 2014.

[FRDC07] F. Fernandez-Riverola, F. Diaz, and J. M. Corchado. Reducing the memory size of a fuzzy case-based reasoning system applying rough set techniques. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(1):138–146, 2007.

[FxYjB+16] M. Fei-xiang, L. Ying-jie, Z. Bo, S. Xiao-yong, and Z. Jing-yu. Intuitionistic fuzzy petri nets for knowledge representation and reasoning. *Journal of Digital Information Management*, 14(2):105, 2016.

[GAG06] S. K. Golmohammadi, A. Azadeh, and A. Gharehgozli. Action selection in robots based on learning fuzzy cognitive map. In *IEEE International Conference on Industrial Informatics*, pages 731–736. IEEE, 2006.

[GAM06] K. M. Gupta, D. W. Aha, and P. Moore. Rough set feature selection algorithms for textual case-based classification. In *European Conference on Case-Based Reasoning*, pages 166–181. Springer, 2006.

[Gre86] J. J. Grefenstette. Optimization of control parameters for genetic algorithms. *IEEE Transactions on systems, man, and cybernetics*, 16(1):122–128, 1986.

[GS08] V. C. Georgopoulos and C. D. Stylios. Complementary case-based reasoning and competitive fuzzy cognitive maps for advanced medical decisions. *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 12(2):191–199, 2008.

[GWL13] N. Ganganath, M. Walker, and H. Leung. Fuzzy cognitive map based situation assessment framework for navigation goal detection. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 1444–1449. IEEE, 2013.

[HDO10] N. Hachani, I. Derbel, and H. Ounelli. Automatic and incremental generation of membership functions. In *Artificial Intelligence and Soft Computing*, pages 97–104. Springer, 2010.

[HL96] T.-P. Hong and C.-Y. Lee. Induction of fuzzy rules and membership functions from training examples. *Fuzzy sets and Systems*, 84(1):33–47, 1996.

[HO07] N. Hachani and H. Ounelli. Improving cluster method quality by validity indices. In *Flairs Conference*, pages 479–483, 2007.

[Hue14] M. Huelsen. *Knowledge-based driver assistance systems: traffic situation description and situation feature relevance*. Springer, 2014.

[Hül07] E. Hüllermeier. *Case-based approximate reasoning*, volume 44. Springer Science & Business Media, 2007.

[HZL+14] C. Huang, S. Zhong, X. Li, F. Zhang, J. Chen, G. Su, Q. Huang, and H. Yuan. Emergency case retrieval based on fuzzy sets and text mining. In *Foundations of Intelligent Systems*, pages 911–919. Springer, 2014.

[JBL06] G. Jakobson, J. Buford, and L. Lewis. A framework of cognitive situation modeling and recognition. In *Military Communications Conference (MILCOM)*, pages 1–7. IEEE, 2006.

[JC11] R. Jensen and C. Cornelis. Fuzzy-rough nearest neighbour classification and prediction. *Theoretical Computer Science*, 412(42):5871–5884, 2011.

[JL95] B. C. Jeng and T.-P. Liang. Fuzzy indexing and retrieval in case-based systems. *Expert Systems with Applications*, 8(1):135–142, 1995.

[JS02] R. Jensen and Q. Shen. Fuzzy-rough sets for descriptive dimensionality reduction. In *IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'02*, pages 29–34. IEEE, 2002.

[JS04] R. Jensen and Q. Shen. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on knowledge and data engineering*, 16(12):1457–1471, 2004.

[JS07] R. Jensen and Q. Shen. Rough set based feature selection: A review. *Rough computing: theories, technologies and applications*, pages 70–107, 2007.

[JSK16] M. Javadian, S. B. Shouraki, and S. S. Kourabbaslou. A novel density-based fuzzy clustering algorithm for low dimensional feature space. *Fuzzy Sets and Systems*, 2016.

[KKC05] R. J. Kuo, Y. P. Kuo, and K.-Y. Chen. Developing a diagnostic system through integration of fuzzy case-based reasoning and fuzzy ant colony system. *Expert Systems with Applications*, 28(4):783–797, 2005.

[KMJS09] A. Khanum, M. Mufti, M. Y. Javed, and M. Z. Shafiq. Fuzzy case-based reasoning for facial expression recognition. *Fuzzy sets and systems*, 160(2):231–250, 2009.

[KML14] J. Kendall-Morwick and D. Leake. A study of two-phase retrieval for process-oriented case-based reasoning. In *Successful Case-based Reasoning Applications-2*, pages 7–27. Springer, 2014.

[Kol14] J. Kolodner. *Case-based reasoning*. Morgan Kaufmann, 2014.

[KPAA14] A. Kofod-Petersen, O. J. Andersen, and A. Aamodt. Case-based reasoning for improving traffic flow in urban intersections. In *International Conference on Case-Based Reasoning*, pages 215–229. Springer, 2014.

[Kun92] L. I. Kuncheva. Fuzzy rough sets: application to feature selection. *Fuzzy sets and Systems*, 51(2):147–153, 1992.

[Lar17] J. Lars. FreeCBR. Available at http://freecbr.sourceforge.net/ (last accessed: 2017/06/28), 2017.

[LBRR07] G. Lavee, A. Borzin, E. Rivlin, and M. Rudzsky. Building petri nets from video event ontologies. *Advances in Visual Computing*, pages 442–451, 2007.

[LH09] S.-T. Li and H.-F. Ho. Predicting financial activity with evolutionary fuzzy case-based reasoning. *Expert Systems with Applications*, 36(1):411–422, 2009.

[LJP14] E. Lupiani, J. M. Juarez, and J. Palma. Evaluating case-base maintenance algorithms. *Knowledge-Based Systems*, 67:180–194, 2014.

[Lok10] S. W. Loke. Incremental awareness and compositionality: A design philosophy for context-aware pervasive systems. *Pervasive and Mobile Computing*, 6(2):239–253, 2010.

[LQL12] G. Lin, Y. Qian, and J. Li. NMGRS: Neighborhood-based multigranulation rough sets. *International Journal of Approximate Reasoning*, 53(7):1080–1093, 2012.

[LTB15] Y. Luo, D. Turgut, and L. Bölöni. Modeling the strategic behavior of drivers for multi-lane highway driving. *Journal of Intelligent Transportation Systems*, 19(1):45–62, 2015.

[LYH12] S. Lee, J. Yang, and J. Han. Development of a decision making system for selection of dental implant abutments based on the fuzzy cognitive map. *Expert Systems with Applications*, 39(14):11564–11575, 2012.

[MAMA14] I. Mala, P. Akhtar, A. R. Memon, and T. J. Ali. Automatic generation of fuzzy membership functions based on k-means and EM clustering. *Sindh University Research Journal (SURJ), Science Series*, 46(2):123–128, 2014.

[MH87] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Readings in nonmonotonic reasoning*, pages 26–45. Morgan Kaufmann Publishers Inc., 1987.

[Min08] Y. Minowa. Verification for generalizability and accuracy of a thinning-trees selection model with the ensemble learning algorithm and the cross-validation method. *Journal of forest research*, 13(5):275–285, 2008.

[MJ12] F. Mohamadnejad and J. Jassbi. A fuzzy inference system platform for simulating cause and effect relationships of strategy map. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2012.

[MLM+05] R. L. D. Mantaras, R. Lopez, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. L. Maher, M. T COX, K. Forbus, et al. Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review*, 20(03):215–240, 2005.

[Mon11] S. Montani. How to use contextual knowledge in medical case-based reasoning systems: A survey on very recent trends. *Artificial intelligence in medicine*, 51(2):125–131, 2011.

[MS05] N. Moray and T. B. Sheridan. on sont les neiges dantan? *Human Performance, Situation Awareness, and Automation: Current Research and Trends HPSAA II, Volumes I and II. Psychology Press*, 2005.

[MSA14] W. Mahdi, M. Soui, and M. Abed. A new personalization approach by case-based reasoning and fuzzy logic. In *International Conference on Advanced Logistics and Transport (ICALT)*, pages 103–108. IEEE, 2014.

[MW16] J. McCulloch and C. Wagner. Measuring the similarity between zSlices general type-2 fuzzy sets with non-normal secondary membership functions. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 461–468. IEEE, 2016.

[Nik08] A. Nikitenko. Knowledge-based robot control. *International Journal ITA*, 15(2):126–137, 2008.

[NKM01] A. Nanopoulos, D. Katsaros, and Y. Manolopoulos. Effective prediction of web-user accesses: A data mining approach. In *Proceedings of the WEBKDD Workshop*, 2001.

[NR03] J. M. Nigro and M. Rombaut. IDRES: A rule-based system for driving situation recognition with uncertainty management. *Information Fusion*, 4(4):309–317, 2003.

[NU09] E. N. Nasibov and G. Ulutagay. Robustness of density-based clustering methods with various neighborhood relations. *Fuzzy Sets and Systems*, 160(24):3601–3615, 2009.

[oG17] University of Girona. eXiTCBR. Available at http://exitcbr.udg.edu/ (last accessed: 2017/06/28), 2017.

[PA15] M. Panda and A. Abraham. Hybrid evolutionary algorithms for classification data mining. *Neural Computing and Applications*, 26(3):507–523, 2015.

[Pap11] E. I. Papageorgiou. A new methodology for decisions in medical informatics using fuzzy cognitive maps based on fuzzy rule-extraction techniques. *Applied Soft Computing*, 11(1):500–513, 2011.

[Paw12] Z. Pawlak. *Rough sets: Theoretical aspects of reasoning about data*, volume 9. Springer Science & Business Media, 2012.

[Per14] P. Perner. Case-based reasoning and the statistical challenges II. In *Man-Machine Interactions 3*, pages 17–38. Springer, 2014.

[PF15] M. S. Pervez and D. M. Farid. Literature review of feature selection for mining tasks. *International Journal of Computer Applications*, 116(21):30–33, 2015.

[PS04] S. K. Pal and S. C. Shiu. *Foundations of soft case-based reasoning*, volume 8. John Wiley & Sons, 2004.

[PS13] E. I. Papageorgiou and J. L. Salmeron. A review of fuzzy cognitive maps research during the last decade. *IEEE Transactions on Fuzzy Systems*, 21(1):66–79, 2013.

[PTH+16] T. Price, M. Tenan, J. Head, W. Maslin, and M. LaFiandra. Acute stress causes over confidence in situation awareness. In *International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 1–6. IEEE, 2016.

[PU15] S. Priyan and R. Uthayakumar. An integrated production–distribution inventory system involving probabilistic defective and errors in quality inspection under variable setup cost. *International Transactions in Operational Research*, 2015.

[PY12] Dillon T. S. Pal, S. K. and D. S. Yeung. *Soft computing in case based reasoning*. Springer Science & Business Media, 2012.

[RADMV09] R. Ros, J. L. Arcos, R. L. De Mantaras, and M. Veloso. A case-based approach for coordinated action selection in robot soccer. *Artificial Intelligence*, 173(9-10):1014–1039, 2009.

[RC12] S. Ranathunga and S. Cranefield. Improving situation awareness in intelligent virtual agents. In *International AAMAS Workshop on Cognitive Agents for Virtual Environments*, pages 134–148. Springer, 2012.

[RFS16] R. Razavi-Far and M. Saif. Imputation of missing data using fuzzy neighborhood density-based clustering. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1834–1841. IEEE, 2016.

[RG02] M. Ruet and L. Geneste. Search and adaptation in a fuzzy object oriented case base. In *European Conference on Case-Based Reasoning*, pages 350–364. Springer, 2002.

[RHM11] D. Ruan, F. Hardeman, and L. Mkrtchyan. Using belief degree-distributed fuzzy cognitive maps in nuclear safety culture assessment. In *Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)*, pages 1–6. IEEE, 2011.

[RJB+14] L. S. Riza, A. Janusz, C. Bergmeir, C. Cornelis, F. Herrera, D. Śle, J. M. Benítez, et al. Implementing algorithms of rough set theory and fuzzy rough set theory in the R package roughsets. *Information Sciences*, 287:68–89, 2014.

[RL14] A. Rodriguez and A. Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

[RMH12] R. M. Rodriguez, L. Martinez, and F. Herrera. Hesitant fuzzy linguistic term sets for decision making. *IEEE Transactions on Fuzzy Systems*, 20(1):109–119, 2012.

[RSDAGC05] J. A. Recio, A. Sánchez, B. Díaz-Agudo, and P. González-Calero. jcolibri 1.0 in a nutshell. a software tool for designing CBR systems. *Proceedings of the 10th UK Workshop on Case Based Reasoning, University of Greenwich; CMS Press*, pages 1–11, 2005.

[SBK16] S. Saha, S. Bhattacharya, and A. Konar. Comparison between type-1 fuzzy membership functions for sign language applications. In *International Conference on Microelectronics, Computing and Communications (MicroCom)*, pages 1–6. IEEE, 2016.

[SE11] A. Smiti and Z. Elouedi. Overview of maintenance for case based reasoning systems. *International Journal of Computer Applications*, 32(2):49–56, 2011.

[Ser10] I. Serina. Kernel functions for case-based planning. *Artificial Intelligence*, 174(16):1369–1406, 2010.

[SGS14] A. Saxena, L. K. Gavel, and M. M. Shrivas. Rough sets for feature selection and classification: An overview with applications. *International Journal of Recent Technology and Engineering (IJRTE)*, 3(5):62–69, 2014.

[Sha92] G. Shafer. Dempster-Shafer theory. *Encyclopedia of artificial intelligence*, pages 330–331, 1992.

[SHGA14] D. H. Sutanto, N. S. Herman, M. Ghani, and K. Abd. Trend of case based reasoning in diagnosing chronic disease: A review. *Advanced Science Letters*, 20(10-11):1740–1744, 2014.

[SHNSon] A. Sarkheyli-Hägele, H. Noguchi, and D. Söffker. Automatic fuzzification of continuous variables for approximate reasoning. journal of approximate reasoning. *Internation Journal of Approximate Reasoning*, 2017, in preparation.

[SHS17a] A. Sarkheyli-Hägele and D. Söffker. Online learning for an individualized lane-change situation recognition system applied to driving assistance. In *IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, pages 1–6. IEEE, 2017.

[SHS17b] Arezoo Sarkheyli-Hägele and Dirk Söffker. Fuzzy SOM-based Case-Based Reasoning for individualized situation recognition applied to supervision of human operators. *Knowledge-Based Systems*, 137:42–53, 2017.

[SHSon] A. Sarkheyli-Hägele and D. Söffker. Learning and representation of event-discrete situations for individualized situation recognition using fuzzy SOM-based CBR. *Engineering Applications of Artificial Intelligence*, 2017, in preparation.

[SHSed] A. Sarkheyli-Hägele and D. Söffker. Integration of case-based reasoning and fuzzy approaches for real-time applications in dynamic environments: Current status and future directions. *Engineering Applications of Artificial Intelligence*, 2017, submitted.

[SJ16] V. K. Singh and R. Jain. *Situation Recognition using EventShop*. Springer, 2016.

[SKK05] R. K. Sharma, D. Kumar, and P. Kumar. FLM to select suitable maintenance strategy in process industries using MISO model. *Journal of Quality in Maintenance Engineering*, 11(4):359–374, 2005.

[SLS11] M. Salamó and M. López-Sánchez. Rough set based approaches to feature selection for case-based reasoning classifiers. *Pattern Recognition Letters*, 32(2):280–292, 2011.

[Smy98] B. Smyth. Case-base maintenance. *Tasks and Methods in Applied Artificial Intelligence*, pages 507–516, 1998.

[Söf01] D. Söffker. From human–machine-interaction modeling to new concepts constructing autonomous systems: A phenomenological engineering-oriented approach. *Journal of Intelligent and Robotic Systems*, 32(2):191–205, 2001.

[Söf08] D. Söffker. Interaction of intelligent and autonomous systems–part i: qualitative structuring of interaction. *Mathematical and Computer modelling of dynamical systems*, 14(4):303–318, 2008.

[Spa01] L. Spalzzi. A survey on case-based planning. *Artificial Intelligence Review*, 16(1):3–36, 2001.

[SS14] B. Sigweni and M. Shepperd. Feature weighting techniques for CBR in software effort estimation studies: a review and empirical evaluation. In *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*, pages 32–41. ACM, 2014.

[SS15] A. Sarkheyli and D. Söffker. Case indexing in case-based reasoning by applying situation operator model as knowledge representation model. *IFAC-PapersOnLine*, 48(1):81–86, 2015.

[SS16a] A. Sarkheyli and D. Söffker. Automatic selection of relevant features using rough set theory for real-time situation recognition based on fuzzy SOM-based CBR. In *International Conference on Systems, Man, and Cybernetics (SMC)*, pages 832–837. IEEE, 2016.

[SS16b] A. Sarkheyli and D. Söffker. Feature selection for situation recognition in Fuzzy SOM-based Case-based Reasoning. In *International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, pages 145–151. IEEE, 2016.

[SSA14] S. V. Shokouhi, P. Skalle, and A. Aamodt. An overview of case-based reasoning applications in drilling engineering. *Artificial Intelligence Review*, pages 1–13, 2014.

[Ste09] W. E. Stempsey. Clinical reasoning: new challenges. *Theoretical medicine and bioethics*, 30(3):173–179, 2009.

[SZ14] E. Sanchez and L. A. Zadeh. *Approximate reasoning in intelligent systems, decision and control*. Elsevier, 2014.

[TBP12] J. Tadrat, V. Boonjing, and P. Pattaraintakorn. A new similarity measure in formal concept analysis for case-based reasoning. *Expert Systems with Applications*, 39(1):967–972, 2012.

[TP09] K. Thangavel and A. Pethalakshmi. Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing*, 9(1):1–12, 2009.

[TSXVV11] G. Tsoumakas, E. Spyromitros-Xioufis, J. Vilcek, and I. Vlahavas. Mulan: A java library for multi-label learning. *Journal of Machine Learning Research*, 12(Jul):2411–2414, 2011.

[Uus07] L. Uusitalo. Advantages and challenges of bayesian networks in environmental modelling. *Ecological modelling*, 203(3):312–318, 2007.

[VDSC15] S. Vluymans, L. Deer, Y. Saeys, and C. Cornelis. Applications of fuzzy rough set theory in machine learning: a survey. *Fundamenta Informaticae*, 142(1-4):53–86, 2015.

[VR08] J. Vascak and M. Rutrich. Path planning in dynamic environment using fuzzy cognitive maps. In *6th International Symposium on Applied Machine Intelligence and Informatics*, pages 5–9. IEEE, 2008.

[VR14] S. Visalakshi and V. Radha. A literature review of feature selection techniques and applications: Review of feature selection in data mining. In *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pages 1–6. IEEE, 2014.

[Wan06] Z. Y. Wang. Developed case-based reasoning system for short-term load forecasting. In *Power Engineering Society General Meeting*, pages 6–pp. IEEE, 2006.

[WB12] A. Wendemuth and S. Biundo. A companion technology for cognitive technical systems. In *Cognitive behavioural systems*, pages 89–103. Springer, 2012.

[WGB08] K. Walzer, M. Groch, and T. Breddin. Time to the rescue-supporting temporal reasoning in the rete algorithm for complex event processing. In *International conference on Database and Expert Systems Applications*, pages 635–642. Springer, 2008.

[WM14] D. Wijayasekara and M. Manic. Data driven fuzzy membership function generation for increased understandability. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 133–140. IEEE, 2014.

[Xio11] N. Xiong. Learning fuzzy rules for similarity assessment in case-based reasoning. *Expert systems with applications*, 38(9):10780–10786, 2011.

[Xio13] N. Xiong. Fuzzy rule-based similarity model enables learning from small case bases. *Applied Soft Computing*, 13(4):2057–2064, 2013.

[XMZ13] N. Xiong, L. Ma, and S. Zhang. Fuzzy relational learning: a new approach to case-based reasoning. In *10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 594–599. IEEE, 2013.

[Yag09] R. R. Yager. Comparing approximate reasoning and probabilistic reasoning using the Dempster–Shafer framework. *International Journal of Approximate Reasoning*, 50(5):812–821, 2009.

[YBG12] N. B. Yahia, N. Bellamine, and H. B. Ghezala. Integrating fuzzy case-based reasoning and particle swarm optimization to support decision making. In *IJCSI: International Journal of Computer Science Issues*, volume 9, pages 117–124. Citeseer, 2012.

[YDM12] J. Ye, S. Dobson, and S. McKeever. Situation identification techniques in pervasive computing: A review. *Pervasive and mobile computing*, 8(1):36–66, 2012.

[YU16] F. G. Yaşar and G. Ulutagay. Challenges and possible solutions to density based clustering. In *IEEE 8th International Conference on Intelligent Systems (IS)*, pages 492–498. IEEE, 2016.

[Zad75] L. A. Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30(3):407–428, 1975.

[ZHQ+15] G.-N. Zhu, J. Hu, J. Qi, J. Ma, and Y.-H. Peng. An integrated feature selection and cluster analysis techniques for case-based reasoning. *Engineering Applications of Artificial Intelligence*, 39:14–22, 2015.

[ZLC14] P. Zhou, S.-W. Lu, and T. Chai. Data-driven soft-sensor modeling for product quality estimation using case-based reasoning and fuzzy-similarity rough sets. *IEEE Transactions on Automation Science and Engineering*, 11(4):992–1003, 2014.

[ZRK11] M. F. Zarandi, Z. S. Razaee, and M. Karbasian. A fuzzy case based reasoning approach to value engineering. *Expert Systems with Applications*, 38(8):9334–9339, 2011.

[ZTCW10] S. Zhao, E. C. Tsang, D. Chen, and X. Wang. Building a rule-based classifiera fuzzy-rough set approach. *IEEE Transactions on Knowledge and Data Engineering*, 22(5):624–638, 2010.

[ZWLP14] J. Zhang, J.-S. Wong, T. Li, and Y. Pan. A comparison of parallel large-scale knowledge acquisition using rough set theory on different MapReduce runtime systems. *International Journal of Approximate Reasoning*, 55(3):896–907, 2014.

This thesis is based on the intermediate results and development steps published/-submitted for publication in the following journals or conferences.

Journal articles

[SHSon] A. Sarkheyli-Hägele and D. Söffker: Learning and representation of event-discrete situations for individualized situation recognition using using fuzzy Situation-Operator Modeling. *Engineering Applications of Artificial Intelligence*, 72:357-367, 2018.

[SHSeda] A. Sarkheyli-Hägele and D. Söffker: Fuzzy SOM-based Case-Based Reasoning for individualized situation recognition applied to supervision of human operators. *Journal of Knowledge-Based System*, 137:42-53, 2017.

[SHSedb] A. Sarkheyli-Hägele and D. Söffker: Integration of Case-Based Reasoning and Fuzzy Approaches for Real-time Applications in Dynamic Environments: Current Status and Future Directions. *Artificial Intelligence Review*, Accepted with minor correction.

[SHNSed] A. Sarkheyli-Hägele, H. Noguchi, and D. Söffker: Automatic fuzzification of continuous variables for approximate reasoning. In preparation.

Conference papers

[SHS17] A. Sarkheyli-Hägele and D. Söffker: Online learning for an individualized lane-change situation recognition system applied to driving assistance. In: *IEEE International Multi-Disciplinary Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, 2017, pages 1-6

[SS16b] A. Sarkheyli and D. Söffker: Feature selection for situation recognition in Fuzzy SOM-based Case-based Reasoning. In: *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 2016, pages 145-151.

[SS16a] A. Sarkheyli and D. Söffker: Automatic selection of relevant features using Rough Set Theory for real-time situation recognition based on fuzzy SOM-based CBR. *IFACPapersOnLine*, 48(2015), pages 81-86.

[SS15] A. Sarkheyli and D. Söffker: Case indexing in Case-Based Reasoning by applying Situation Operator Model as knowledge representation model. In: *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2015, pages 832-837.

In the context of research projects at the Chair of Dynamics and Control, the following student theses have been supervised by Arezoo Sarkheyli-Hägele and Univ.-Prof. Dr.-Ing. Dirk Söffker. Development steps and results of the research projects and the student theses are integrated with each other and hence are also part of this thesis.

[Nog16] H. Noguchi. Automatic generation of fuzzy membership functions for dynamic environment variables. Master Thesis, October 2016.

[Xu16] M. Xu. Driving situation recognition using classification methods for driving assistance system. Master Thesis, November 2016.

Additionally, in the context of research projects at the Chair of Dynamics and Control, the following student theses have been supervised by Arezoo Sarkheyli-Hägele and Univ.-Prof. Dr.-Ing. Dirk Söffker.

[Che16] L. Chen. Intention recognition of human driver using Bayesian Networks. Bachelor Thesis, September 2015.

[Ji16] Q. Ji. Development of a filtering and processing tool for driving simulator and eye tracker data. Bachelor Thesis, March 2016.

[Mar16] M. F. Marques. Erstellung eines Szenarios zur Analyse und Evaluierung von Assistenz- und Entscheidungsunterstützungssytemen. Bachelor Thesis, July 2016.