**IMT Institute for Advanced Studies, Lucca**

Lucca, Italy

# Automatic Context Adaptation of Fuzzy Systems

PhD Program in Computer Science and Engineering

XX Cycle

**By**

**Alessio Botta**

**2008**

**The dissertation of Alessio Botta is approved.**

Program Coordinator: Prof. Ugo Montanari, University of Pisa

Supervisor: Prof. Beatrice Lazzerini, University of Pisa

Supervisor: Prof. Francesco Marcelloni, University of Pisa

Tutor: Prof. Beatrice Lazzerini, University of Pisa

The dissertation of Alessio Botta has been reviewed by:

Prof. Oscar Cordón, European Centre for Soft Computing, Mieres, Spain

Prof. Witold Pedrycz, University of Alberta, Edmonton, AB, Canada

**IMT Institute for Advanced Studies, Lucca**

**2008**

To my family

# Contents

# List of Figures

# List of Tables

# Acknowledgements

I would like to acknowledge my supervisors, Prof. Beatrice Lazzerini and Prof. Francesco Marcelloni, University of Pisa, who advised me during the three years of my PhD studies, giving me full support in the development of both my research and my thesis.

I would also like to thank Prof. Dan C. Stefanescu, Suffolk University, who contributed to a relevant part of the research on context adaptation of fuzzy systems and who generously spent his time and efforts to organize my visiting period in Boston.

I am very grateful to Prof. Witold Pedrycz, University of Alberta, and to Prof. Oscar Cordón, European Centre for Soft Computing, who kindly agreed to review my thesis and dedicated their time to provide deep feedbacks on the current work and valuable ideas for future extensions.

Silvia Lucchesi, IMT Lucca Institute for Advanced Studies, has given me help and tutorship for the all the paperwork and administrative stuff. I thank Silvia for her infinite patience and friendship.

Finally, I would like to acknowledge my fellows Alessandro Erta, Michele Zanda, Davide Bacciu, Antonio Bucchiarone, and Leonardo Badia, IMT Lucca Institute for Advanced Studies, with whom I shared my working and leisure time during the last three years. The several discussions on topics not necessarily related to my research area and the advices that we shared gave me a number of stimuli to achieve better results and to keep my mind open to different subjects.

# Vita

| | |
|---|---|
| **August 10, 1980** | Born, Livorno, Italy |
| **2002** | Laurea Degree in Information Technology Engineering<br>Final mark: 110/110 cum laude<br>University of Pisa, Pisa, Italy |
| **2004** | Laurea Specialistica Degree in Information Technology Engineering for Business Management<br>Final mark: 110/110 cum laude<br>University of Pisa, Pisa, Italy |
| **2006** | Senior Lecturer and Research Scholar<br>Maths & Computer Science Department<br>Suffolk University, Boston (MA), United States |
| **2007** | Teaching Assistant<br>Department of Information Engineering<br>University of Pisa, Pisa, Italy |

# Publications


1. A. Botta, P. Ducange, B. Lazzerini, and F. Marcelloni. A multi-objective cooperative coevolutionary approach to Mamdani fuzzy system generation. In *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2008)*, Malaga, Spain, 2008.
2. A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, to appear, 2008.
3. A. Botta, B. Lazzerini, and F. Marcelloni. Context adaptation of Mamdani fuzzy rule-based systems. *International Journal of Intelligent Systems*, 23(4):397–418, 2008.
4. D. Bacciu, A. Botta, and L. Badia. Fuzzy admission control with similarity evaluation for VoWLAN with QoS support. In *Proceedings of the 5th Annual Conference on Wireless On demand Network Systems and Services (IEEE/IFIP WONS 2008)*, Garmisch-Partenkirchen, Germany, 2008.
5. A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. A survey of approaches to context adaptation of fuzzy systems. In *Proceedings of the 10th IASTED International Conference on Intelligent Systems and Control (ISC 2007)*, Cambridge, MA, USA, 2007.
6. D. Bacciu, A. Botta, and D. C. Stefanescu. A framework for semantic querying of distributed data-graphs via information granules. In *Proceedings of the 10th IASTED International Conference on Intelligent Systems and Control (ISC 2007)*, Cambridge, MA, USA, 2007.
7. A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Exploiting fuzzy ordering relations to preserve interpretability in context adaptation of fuzzy systems. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'07)*, pages 1–6, London, UK, 2007.
8. L. Badia, A. Botta, and L. Lenzini. Joint routing and link scheduling for wireless mesh networks through genetic algorithms. In *Proceedings of the 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt 2007)*, Lymassol, Cyprus, 2007.
9. D. Bacciu, A. Botta, and H. Melgratti. A fuzzy approach for negotiating quality of services. In *Proceedings of the 2nd Symposium on Trustworthy Global Computing (TCG 2006)*, volume 4661 of *Lecture Notes in Computer Sciences*, pages 200–217, Lucca, Italy, 2007.

10. D. Bacciu, A. Botta, and D. C. Stefanescu. Augmenting the distributed evaluation of path queries with information granules. In *Proceedings of the 5th International Workshop on Mining and Learning with Graphs (MLG'07)*, pages 105–109, Firenze, Italy, 2007.

11. D. Bacciu, L. Badia, and A. Botta. Fuzzy agreement for network service contracts. In *Proceedings of the 10th Joint Conference on Information Sciences*, pages 349–355, Salt Lake City, UT, USA, 2007.

12. A. Botta, B. Lazzerini, and F. Marcelloni. New operators for context adaptation of Mamdani fuzzy systems. In *Applied Artificial Intelligence: Proceedings of the 7th Internations FLINS Conference*, D. Ruan, P. D'hondt, P. F. Fantoni, M. D. Cock, M. Nachtegael, and E. E. Kerre, Eds., World Scientific, pages 35–42, Genova, Italy, 2006.

13. A. Botta, B. Lazzerini, and F. Marcelloni. Context adaptation of Mamdani fuzzy systems through new operators tuned by a genetic algorithm. In *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'06)*, pages 1641–1648, Vancouver, BC, Canada, 2006.

14. A. Botta, B. Lazzerini, and F. Marcelloni. Integrating service-oriented technologies to support business processes. In *Proceedings of the 7th IEEE International Conference on E-Commerce Technology Workshops (SoS4CO)*, pages 37–42, Munich, Germany, 2005.

# Presentations

1. A. Botta, “Evolutionary algorithm techniques for network optimization problems”. *Colloquia of the Department of Information Engineering, University of Padova*, Padova, Italy, 25/10/2007.
2. A. Botta, “Augmenting the distributed evaluation of path queries with information granules”. *5th International Workshop on Mining and Learning with Graphs (MLG’07)*, Firenze, Italy, 01/08/2007.
3. A. Botta, “Exploiting fuzzy ordering relations to preserve interpretability in context adaptation of fuzzy systems”. *2007 International Conference on Fuzzy Systems (FUZZ-IEEE’07)*, London, UK, 25/07/2007.
4. A. Botta, “Computational intelligence and applications”. *Summer School of Information Engineering*, Bressanone, Italy, 04/07/2007.
5. A. Botta, “Automatic email classification: An application-oriented review”. *IMT Lucca Institute for Advanced Studies Seminars*, Lucca, Italy, 11/04/2006.
6. A. Botta, “Integrating service-oriented technologies to support business processes”. *The First IEEE International Workshop on Service Oriented Solutions for Cooperative Organizations (SoS4CO)*, Munich, Germany, 19/07/2005.

# Abstract

Among the several applications of fuzzy set theory, fuzzy-rule based systems (FRBSs) have proven to be extremely successful in a wide range of fields, including, for instance, control, classification, regression, and pattern recognition. In particular, FRBSs have raised attention for their twofold nature, i.e., for their ability to handle linguistic concepts and, at the same time, to perform an accurate modeling of input-output relations. Hence, several researchers and practitioners have developed learning algorithms for the automatic identification of FRBSs from real-world data. Such algorithms include the hybridizations of FRBSs with other popular soft computing techniques, i.e., artificial neural networks and evolutionary algorithms. In this framework, context adaptation of fuzzy systems is considered as an emerging paradigm which has been analyzed only in a few works. In a nutshell, context adaptation consists in tuning some of the features of an already existing FRBS, so as to adapt it to a new configuration of the external environment. This task has usually been approached in the literature as scaling fuzzy sets from a universe of discourse to another. The basic idea presented here is to achieve context adaptation by exploiting a set of operators which allow performing a more flexible tuning than scaling-function-based techniques, while keeping the semantics and the interpretability of the original FRBS unaltered. Nonetheless, this work collects previous approaches and organizes them in a common framework, thus providing a reference study on the topic. In the development of the thesis, we first recall fundamentals of fuzzy set theory, fuzzy logic, and fuzzy systems. We survey previous work on related subjects and introduce

the context adaptation problem in detail. Second, we develop a novel context adaptation approach. To this aim, we introduce a flexible non linear scaling function and four orthogonal fuzzy modifiers which allow adapting an FRBS to any context. Since the modeling capabilities of the operators may negatively affect the semantics of the FRBS, we study two novel indices to properly measure interpretability and prevent such degradation. The proposed learning approach is based on evolutionary algorithms and takes both the flexibility introduced by the operators and the interpretability assessed by the indices into account. We test our context adaptation technique on four different data sets, providing detailed examples and comparisons. Finally, we draw concluding remarks and we discuss future extensions and possible research lines.

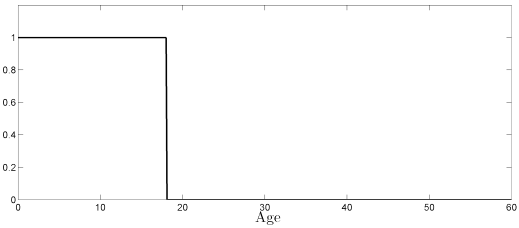
# Chapter 1

# Introduction

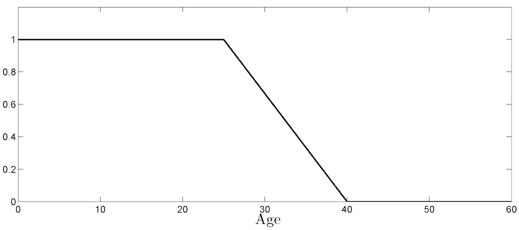
*Fuzzy set theory* (FST) was introduced by Zadeh in his seminal work *Fuzzy sets*, published on *Information and Control* in 1965 (Zad65). The original idea that lies behind fuzzy sets is that, while some concepts are easily represented by classical set theory, other are difficult to express within such framework, because they intrinsically involve some degrees of uncertainty and vagueness.

For instance, consider the set comprising the "*people under 18 years in Lucca*" and the one comprising "*young people in Lucca*". The former set can be easily and exclusively identified by computing the age of all the inhabitants of Lucca, as in Figure 1(a), while the latter is quite imprecisely defined and difficult to identify with traditional mathematical tools. To take such uncertainty and vagueness into account, Zadeh introduced the notion of *fuzzy set*.

Typically, the characteristic function of a set has only two possible values: zero, if the element does not belong to that set, or one, if the element is part of the set. Zadeh's idea is to allow the values of the characteristic function varying in the continuous range $[0, 1]$, thus giving to the set a much more effective expressive power in representing imprecise concepts. Indeed, using a fuzzy set, we may define the set of "*young people in Lucca*" as in Figure 1(b), i.e., giving the maximum degree of membership of 1 to people with age under 25, and decreasing the degree of member-




(a)




(b)

**Figure 1:** (a) Characteristic function of the set "*people under 18 in Lucca*" and (b) a possible MF for the fuzzy set "*young people in Lucca*"

ship until reaching 0 for people over 40. This can be linguistically interpreted as "*people under 25 are young to the maximum degree, people from 25 to 40 are still young but not completely, people over 40 are not young anymore*". Such real-valued characteristic functions are known as *membership functions* (MFs).

Extending the idea of fuzzy membership to logic is straight forward: rather than using just *false* and *true* as truth degrees, we consider a whole set of continuous truth values ranging in the interval $[0, 1]$. *Fuzzy logic* (FL), differently from classical two-valued logics, aims to model the imprecise way of reasoning of human brain and its ability to make rational decisions based on uncertainty and imprecision. FL is often related to *computing with words* by its creator (Zad96), because, in fuzzy inference systems, fuzzy logic is commonly employed to perform computation over *linguistic variables*.

FST and FL have proven to be very effective in commercial and industrial applications, including biomedical engineering, robotics, data analysis and regression, pattern recognition, image processing, and, more in general, in several kinds of filtering, modeling and control applications (KY95; KY96; BKKP99; DLJ00; Ros04).

We remark that FST and FL are building blocks of the wider framework of *soft computing* (SC), a collection of techniques designed to handle extremely hard problems in which the application of traditional approaches fails. As stated by Zadeh (Zad94),

> Soft computing is not a homogeneous body of concepts and techniques. Rather, it is a partnership of distinct methods that in one way or another conform to its guiding principle. At this juncture, the dominant aim of SC is to exploit the *tolerance for imprecision and uncertainty* to achieve tractability, robustness, and low solution cost. The principal constituents of SC are fuzzy logic, neurocomputing, and probabilistic reasoning, with the latter subsuming genetic algorithms, belief networks, chaotic systems, and parts of learning theory. [...] In large measure, fuzzy logic, neurocomputing, and probabilistic reasoning are complementary, not competitive. [...] It is advantageous to combine them.

A number of techniques have been developed to ease the development of systems based on fuzzy techniques from real-world data, including hybrid systems built upon the combination with other SC techniques such as *artificial neural networks* (Jan93; Hay94; Abr01) and *evolutionary algorithms* (Mic99; CGH$^+$04; Her08).

Getting back to the example on young people, it is worth noting that the concept of "*young people in Lucca*" is not only intrinsically fuzzy and imprecise, but, also, it cannot be uniquely and universally defined. Indeed, different people could define different fuzzy sets to represent the same shared concept of *youth*. To better illustrate this idea, let us refer to a clear example provided by Cordón et al.

In (CHMV01), the authors state that people tend to set the upper limit of the support of the fuzzy set *young* not at any fixed value, like 40 in the previous example, but rather to a value equal to $own\_age + 5$. In other words, the exact definition of the fuzzy set depends on an external factor (i.e., $own\_age$) which is strongly user-dependent. Although funny, this is a typical example of *context adaptation* of a fuzzy set.

Basically, context adaptation of fuzzy systems refers to a set of techniques that are used to generate specialized instances of fuzzy systems from generic models. In the past fifteen years, context adaptation has been considered a promising approach to the development of algorithms for the automatic identification of fuzzy systems from data, but it has never been deeply explored.

Hence, the primary objective of this thesis is to serve as a reference for the topic of context adaptation of fuzzy system. To this aim, we collected and surveyed previous work and provided a general framework for context adaptation which is coherent with existing approaches. Further, we introduced novel operators and studied interpretability issues related to context adaptation. Finally, we exploited evolutionary algorithms to perform automatic adaptation from real-world data.

The rest of the thesis describes our research in detail and is organized as follows.

- Chapter 2 (**Backgrounds and Previous Work**) is divided in two main Sections. In Section 2.1, we recall fundamentals and references of FST, FL, and fuzzy systems. In Section 2.2, we first introduce the concept of context and its relation with FST. Then, we survey previous work on related topics and, finally, we develop a common framework that serves as a reference for the following Chapters. Part of the work of Section 2.2 has been published in (BLMS07b).
- In Chapter 3 (**Novel Operators for Context Adaptation of Fuzzy Partitions**), we introduce some novel operators that have been carefully designed to perform a flexible context adaptation of fuzzy sets. For each operator, we provide a detailed analysis and sample applications. Preliminary versions of these operators have been introduced in (BLM06a; BLM06b; BLMS08; BLM08).

- In Chapter 4 (**Interpretability Issues in Context Adaptation of Fuzzy Systems**), we deal with the accuracy-interpretability trade-off of fuzzy systems. We analyze the effects of a flexible context adaptation on the interpretability of a fuzzy partition and we develop two alternative approaches to provide countermeasures. Part of the work of Chapter 4 has been published in (BLMS07a; BLM08; BLMS08; BDLM08).

- In Chapter 5 (**Evolutionary Algorithms for Automatic Context Adaptation of Fuzzy Systems**), we introduce three evolutionary algorithms that, exploiting the operators introduced in Chapter 3 and the approaches to preserve interpretability developed in Chapter 4, can be used for the identification of context-adapted genetic fuzzy systems. Preliminary versions of the proposed algorithms have been developed in (BLM06a; BLM06b; BLMS07a; BLM08; BLMS08).

- In Chapter 6 (**Experimental Results**), we assess the proposed operators and algorithms on four different context-aware data sets.

- Finally, in Chapter 7 (**Conclusion and Future Work**), we draw final conclusions. Moreover, we analyze possible extensions of the current work and we hypothesise future research trends on context adaptation of fuzzy systems.

# Chapter 2

# Backgrounds and Previous Work

In this Chapter, we recall fundamentals and references on fuzzy set theory, fuzzy logic, and fuzzy systems. Further, we survey previous work on context adaptation of fuzzy systems and provide a general introduction on the topic.

## 2.1 Backgrounds

FST and FL are well-estabilished topics that have been studied and applied for decades. In this Section, we provide a short reference of typical definitions and concepts that will be particularly useful in the rest of the thesis.

The interested reader can find several authoritative resources related to backgrounds of FST and FL, for instance in (Zad65; Zad73; Mam76; Zad94; KY95; KY96; DLJ00; CS02; CCHM03; Ros04).

### 2.1.1 Fundamentals of Fuzzy Set Theory

This Section introduces some common definitions about FST that will be exploited in the following Chapters. Similar definitions can be found in (dO99; DLJ00; Coc01; CK02).

**Definition 1 (Fuzzy set)** *Let $\mathcal{U}$ be a space of points (objects), where a generic element of $\mathcal{U}$ is denoted by $x$. Thus, $\mathcal{U} = \{x\}$. A fuzzy set $A$ in $\mathcal{U}$ is characterized by a membership function $A(x) : \mathcal{U} \rightarrow [0, 1]$, where $[0, 1] \subset \mathbb{R}$ and $\mu_A(\hat{x})$ denotes the degree of membership to which $\hat{x}$ belongs to $A$.*

For the sake of readability, we will often use the universally accepted notation that identifies a fuzzy set with its MF, i.e., we will denote $\mu_A(x)$ as $A(x)$. It can be observed that the definition of fuzzy set extends the traditional definition of set: in the following, we will refer to traditional sets as *crisp sets*.

**Definition 2 (Crisp set)** *A crisp set $A$ is a fuzzy set where $\forall x \in \mathcal{U}$, $A(x) = 0 \vee A(x) = 1$.*

Two useful concepts that will be exploited in Chapter 3 are those of *support* and *core* of a fuzzy set.

**Definition 3 (Support of a fuzzy set)** *The support of a fuzzy set $A$ is the crisp set $\mathcal{S}(A) = \{x \in \mathcal{U} | A(x) > 0\}$.*

**Definition 4 (Core of a fuzzy set)** *The core of a fuzzy set $A$ is the crisp set $\mathcal{C}(A) = \{x \in \mathcal{U} | A(x) = 1\}$.*

FST extends the basic concepts and operators of equivalence, complement, subsetting, intersection and union defined in traditional set theory. To define FST operators, we first have to introduce the concepts of *t-norm* and *t-conorm*.

**Definition 5 (T-norm)** *A t-norm $t$ is a function $t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that, $\forall a, b, d \in [0, 1]$, (i) $t(a, 1) = a$, (ii) $b \leq d \Rightarrow t(a, b) \leq t(a, d)$, (iii) $t(a, b) = t(b, a)$, and (iv) $t(a, t(b, d)) = t(t(a, b), d)$. A t-norm is usually continuous and such that $t(a, a) \leq a \; \forall a \in [0, 1]$.*

**Definition 6 (T-conorm)** *A t-conorm $s$ is a function $s : [0, 1] \times [0, 1] \rightarrow [0, 1]$ such that, $\forall a, b, d \in [0, 1]$, (i) $s(a, 0) = a$, (ii) $b \leq d \Rightarrow s(a, b) \leq s(a, d)$, (iii) $s(a, b) = s(b, a)$, and (iv) $s(a, s(b, d)) = s(s(a, b), d)$. A t-conorm is usually continuous and such that $s(a, a) \geq a \; \forall a \in [0, 1]$.*

We remark that a t-conorm can be defined in terms of a t-norm, as

$$s(a, b) = 1 - t(1 - a, 1 - b), \; \forall a, b \in [0, 1]. \tag{2.1}$$

**Definition 7 (Empty fuzzy set)** *A fuzzy set $A$ is empty iff $\forall x \in \mathcal{U}$, $A(x) = 0$.*

**Definition 8 (Equivalence of fuzzy sets)** *Two fuzzy sets $A$ and $B$ in $\mathcal{U}$ are equal iff $\forall x \in \mathcal{U}$, $A(x) = B(x)$.*

**Definition 9 (Fuzzy subset)** *Given two fuzzy sets $A$ and $B$, $A \subseteq B$ holds iff $\forall x \in \mathcal{U}$, $A(x) \leq B(x)$.*

**Definition 10 ($\alpha$-cut)** *Given $\alpha \in [0, 1]$, the $\alpha$-cut $A_\alpha$ of a fuzzy set $A$ is the crisp set $A_\alpha = \{x | A(x) \geq \alpha\}$.*

Given Definitions 2 – 4 and 9, we observe that the support and the core of a fuzzy set $A$ can be interpreted as the smallest crisp superset and the largest crisp subset of $A$, respectively. Also, we remark that, in light of Definition 10, $\mathcal{C}(A)$ and $\mathcal{S}(A)$ are the 1-cut and the $0^+$-cut of $A$, respectively.

**Definition 11 (Complement of a fuzzy set)** *The complement $\bar{A}$ of a fuzzy set $A$ is characterized by the MF $\bar{A}(x) = 1 - A(x)$.*

**Definition 12 (Intersection of fuzzy sets)** *Given two fuzzy sets $A$ and $B$, their intersection $C = A \cap B$ is characterized by the MF $C(x) = t\left(A\left(x\right), B\left(x\right)\right)$, where $t(a, b)$ is a t-norm.*

**Definition 13 (Union of fuzzy sets)** *Given two fuzzy sets $A$ and $B$, their union $C = A \cup B$ is characterized by the MF $C(x) = s\left(A\left(x\right), B\left(x\right)\right)$, where $s(a, b)$ is a t-conorm.*

In the rest of the thesis, the $\min$ and the $\max$ operators are used as t-norm and t-conorm to compute intersection and union of fuzzy sets, respectively. Indeed, these functions are widely accepted as standard FST operators in the literature.

The following definitions will be useful to measure and classify fuzzy sets in the rest of the thesis.

**Definition 14 (Cardinality of a fuzzy set)** *The cardinality of a fuzzy set $A$ is given by $|A| = \displaystyle\int_{\mathcal{U}} A(x)\, dx$.*

**Definition 15 (Height of a fuzzy set)** *The height of a fuzzy set $A$ is given by $h(A) = \sup\limits_{x \in \mathcal{U}} A(x)$.*

**Definition 16 (Normal fuzzy set)** *A fuzzy set $A$ is normal iff $h(A) = 1$, subnormal iff $h(A) < 1$.*

**Definition 17 (Convex fuzzy set)** *A fuzzy set $A$ is convex iff $\forall l, r, x \in X$, $l \leq x \leq r \Rightarrow A(x) \geq \min(A(l), A(r))$.*

**Definition 18 (Unimodal fuzzy set)** *A fuzzy set $A$ is unimodal iff it is normal and its core $\mathcal{C}(A)$ is a convex set.*

**Definition 19 (Continuous fuzzy set)** *A fuzzy set $A$ is continuous iff its MF is continuous in $\mathcal{U}$.*

**Definition 20 (Fuzzy number)** *A fuzzy number is a normal and convex fuzzy set defined on the universe $\mathbb{R}$ of real numbers.*

We remark that fuzzy numbers are usually defined on a bounded universe $\mathcal{U} = [u_{\min}, u_{\max}] \subset \mathbb{R}$. The following concepts related to *fuzzy modifiers* will be of particular interest in Chapter 3. Some of these definitions are taken by (CK00; Coc01; CK02).

**Definition 21 (Fuzzy modifier)** *Let $\mathcal{F}(\mathcal{U})$ be the set of all fuzzy sets on $\mathcal{U}$. A fuzzy modifier is any general mapping $m : \mathcal{F}(\mathcal{U}) \rightarrow \mathcal{F}(\mathcal{U})$.*

**Definition 22 (Expansive fuzzy modifier)** *A fuzzy modifier $m$ is expansive iff $\forall A \in \mathcal{F}(\mathcal{U})$, $A \subseteq m(A)$.*

**Definition 23 (Restrictive fuzzy modifier)** *A fuzzy modifier $m$ is restrictive iff $\forall A \in \mathcal{F}(\mathcal{U})$, $A \supseteq m(A)$.*

**Definition 24 (Inclusive fuzzy modifier)** *A fuzzy modifier is inclusive if it is either expansive or restrictive.*

**Definition 25 (Pre- and post-modifier)** *A fuzzy modifier $m$ in $\mathcal{U}$ is decomposable in a pre- and a post-modifier iff there exists a mapping $p : \mathcal{U} \rightarrow \mathcal{U}$ and a mapping $q : [0,1] \rightarrow [0,1]$ such that $\forall A \in \mathcal{F}(\mathcal{U})$, $m(A(x)) = q(A(p(x)))$. Relations $p$ and $q$ are called the pre-modifier and the post-modifier of $m$, respectively.*

Definition 25 can be detailed as follows.

**Definition 26 (Pure pre-modifier)** *A fuzzy modifier $m$ is said to be a pure pre-modifier iff it is decomposable and $q$ is the identity relation, i.e. iff $m(A(x)) = A(p(x))$.*

**Definition 27 (Pure post-modifier)** *A fuzzy modifier $m$ is said to be a pure post-modifier iff it is decomposable and $p$ is the identity relation, i.e. iff $m(A(x)) = q(A(x))$.*

Finally, we define the concept of *fuzzy relation*, which will be exploited in the following Section to define fundamentals of fuzzy logic and in Chapter 4 to derive an interpretability index based on fuzzy ordering relations.

**Definition 28 (Fuzzy relation)** *Given a set of crisp sets $\mathcal{U}_1, ..., \mathcal{U}_n$, a fuzzy relation $R : \mathcal{U}_1 \times ... \times \mathcal{U}_n \to [0, 1]$ is a fuzzy set defined on the Cartesian product of crisp sets $\mathcal{U}_1, ..., \mathcal{U}_n$, where tuples belonging to the Cartesian product may have varying degrees of membership within the relation.*

### 2.1.2 Fundamentals of Fuzzy Logic

The idea of a multi-valued logic was not completely new at the time Zadeh introduced it (Bru92). Indeed, Plato, Heraclitus and other ancient Greek philosophers had already noted that the *law of the excluded middle*, which states that every proposition must either be true or false, is somehow faulty.

At the beginning of the 20th century, Łukasiewicz introduced a multi-valued logic with a third unknown value, together with new axioms and an alternative logic theory (Bor70). After that first attempt, a number of multi-valued logics have been developed: among them all, the infinite-valued logic introduced by Zadeh is by far the most well-known and applied. The interested reader can find an historical perspective of fuzzy logic in (Ros04).

Fuzzy logic is isomorphic to fuzzy set theory in the same way as traditional two-valued logic is isomorphic to crisp set theory (KY95). Roughly speaking, the isomorphism follows from the fact that the logic operations have the same mathematical form as the corresponding operations on fuzzy sets. Since fuzzy operators can be implemented in different ways, a variety of fuzzy set theories and, therefore, of fuzzy logics can be derived.

This Chapter introduces the concepts of fuzzy logic that are exploited to develop the fuzzy systems which will be shown in Section 2.1.3.

**Definition 29 (Linguistic variable)** *A linguistic variable $X$ is a variable which takes linguistic terms as values. A linguistic variable is defined by a quintuple $(X, \mathcal{T}(X), \mathcal{U}, R, M)$, where $X$ is the name of the variable (e.g., Age), $\mathcal{T}(X)$ is the set of linguistic terms of $X$ (e.g., $\mathcal{T}(X) = \{young,old\}$), $\mathcal{U}$ is the universe of discourse of the base variable, $R$ is a syntactic rule for generating composed linguistic terms of $X$ (e.g., very cold or not young), and $M : \mathcal{T}(X) \to \mathcal{F}(\mathcal{U})$ is a semantic rule mapping each element in $\mathcal{T}(X)$ with a fuzzy set defined over the universe of discourse $\mathcal{U}$ of the base variable.*

Most of the times, fuzzy numbers are used as meanings of linguistic terms. Once we have defined linguistic variables, it is possible to derive an *approximate reasoning* process, by introducing fuzzy logic connectives and inference rules.

**Definition 30 (Fuzzy proposition)** *A proposition $P$ in fuzzy logic takes the form $P$ : $x$ is $A$, where $x$ is a variable defined on the universe $\mathcal{U}$ and $A$ is a linguistic term bound to a fuzzy set $A(x) \in \mathcal{F}(\mathcal{U})$. The degree of truth of $P$ is $T(P) = A(x)$.*

**Definition 31 (Negation)** *Given a proposition $P$ : $x$ is $A$, the degree of truth of $\neg P$ is $T(\neg P) = \bar{A}(x) = 1 - A(x)$.*

**Definition 32 (Conjunction)** *Given two propositions $P$ : $x$ is $A$ and $Q$ : $y$ is $B$, the degree of truth of $P \wedge Q$ is $T(P \wedge Q) = t(A(x), B(y))$, where $t(a, b)$ is a t-norm.*

**Definition 33 (Disjunction)** *Given two propositions $P$ : $x$ is $A$ and $Q$ : $y$ is $B$, the degree of truth of $P \vee Q$ is $T(P \vee Q) = s(A(x), B(y))$, where $s(a, b)$ is a t-conorm.*

**Definition 34 (Implication)** *Given two propositions $P$ : $x$ is $A$ and $Q$ : $y$ is $B$, the degree of truth of $P \to Q$ is $T(P \to Q) = I(A(x), B(y))$, where $I : [0, 1] \times [0, 1] \to [0, 1]$ is any fuzzy relation which extends boundary conditions of classical implication, i.e. $I(0, 0) = I(0, 1) = I(1, 1) = 1$ and $I(1, 0) = 0$.*

As in the previous Section, we refer to $\min$ and $\max$ as standard t-norm and t-conorm to implement the conjunction and disjunction logical operators, respectively. On the contrary, there is no standard definition of implication on which authors widely agreed. Surprisingly, one of the most popular choice for fuzzy implication, which has been introduced by

Mamdani in (Mam76), does not comply with all of the four constraints. Indeed, Mamdani proposed the use of the $\min$ function also to implement implication, i.e.

$$I\left(A\left(x\right),B\left(y\right)\right)=\min\left(A\left(x\right),B\left(y\right)\right). \tag{2.2}$$

Trivially, $\min(0,1) = 0$ and, therefore, the boundary conditions are not satisfied. Nonetheless, the cases in which we have $A(x) = 0$ and $B(y) = 1$ are pretty exceptional in real-world applications. Moreover, Mamdani's implication is computationally cheap and, therefore, it is widely applied to fuzzy systems. In the rest of this thesis, we will consider Mamdani's implication as the standard form of implication.

The essential concept of approximate reasoning is that, given an inference rule and an uncertain value for the premise, a new consequent can be derived. This can be achieved by several different inference engines, including the well-known *generalized modus ponens*.

**Definition 35 (Generalized modus ponens)** *Given a rule*

$$r: \text{ if } x \text{ is } A \text{ then } y \text{ is } B,$$

*and a premise*

$$x \text{ is } A',$$

*where $x \in \mathcal{U}$, $y \in \mathcal{V}$, $A, A' \in \mathcal{F}(\mathcal{U})$, and $B \in \mathcal{F}(\mathcal{V})$, we can derive a new consequent $B'$ via the generalized modus ponens as*

$$B'(y)=\sup_{x\in\mathcal{U}} t\left(A'\left(x\right),I\left(A\left(x\right),B\left(y\right)\right)\right), \tag{2.3}$$

*where t is a t-norm and $I$ is a fuzzy implication.*

Obviously, the generalized modus ponens can be implemented in several different ways, depending on the choice of the t-norm and of the implication. As stated above, we make use of the $\min$ operator to implement both $t$ and $I$.

### 2.1.3 Fuzzy Systems

Generically speaking, any system based on FST and/or FL techniques may be considered as a fuzzy system. Hence, fuzzy systems may include, for instance, clustering applications (DLJ00; XI05; CLM06), filters

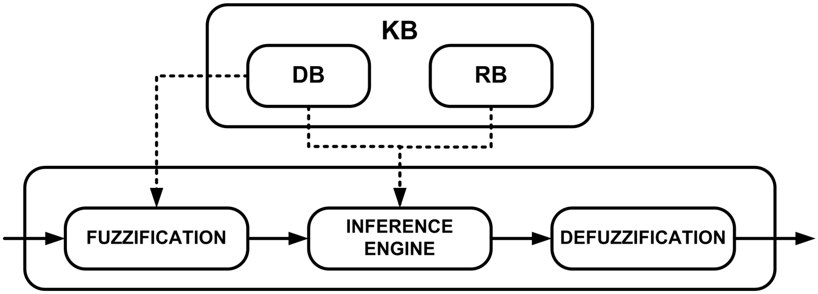



**Figure 2:** The architecture of a Mamdani-type FRBS

for image processing (BKKP99), linguistic models (PV99), fuzzy-logic-enabled data bases (BKP05), service agreement frameworks (BBM07), etc. Nonetheless, the term fuzzy system is commonly employed to refer to a given class of expert systems characterized by a linguistic rule base and by an FL-based inference engine. To avoid confusion, in the following we will often refer to such expert systems as *fuzzy rule based systems* (FRBSs), or *fuzzy inference systems* (FISs).

Due to the increasing popularity of FST-based techniques, different kinds of FRBSs have been developed to cope with several applications, including FRBSs for classification, regression, and control. Specifically, we refer to Mamdani-type FRBSs (MA75), which are, by and large, recognized as the most transparent (i.e., human-interpretable) kind of FRBSs.

Formally, a Mamdani-type FRBS is a mathematical model that, given $n$ inputs $x_1, ..., x_n$, computes an output $y$, exploiting the knowledge coded in a *rule base* (RB) and in a *data base* (DB), and an inference process based on fuzzy logic.

The RB is composed of $R$ linguistic rules in the form

$$r: \text{ if } X_1 \text{ is } A_1^r(x_1) \text{ and ... and } X_n \text{ is } A_n^r(x_n) \text{ then } Y \text{ is } B^r(y). \tag{2.4}$$

A linguistic variable $X_i$ is defined for each input $x_i$, and, therefore, $A_i^r(x_i)$ are the fuzzy sets associated with the linguistic terms of $X_i$ by the semantic rule $M$, as explained in Definition 29. In the following, we will blur the distinction between linguistic terms and associated fuzzy sets. Similarly, a linguistic variable $Y$ is also defined for the output $y$. The knowledge represented by the linguistic variables is comprised in the DB. The DB and the RB are globally referred to as the *knowledge base* (KB). The output $y$ is computed by the FL-based inference engine, which, in a nutshell, consists in the aggregation of the consequents of all the rules that match with the inputs. The overall structure of a Mamdani-type FRBS is depicted in Figure 2.

The steps performed by the inference engine to compute the output are the following.

1. *Fuzzification*: compare the crisp input variables with the MFs of the antecedent part of each rule, so as to compute the membership degrees of the inputs for each linguistic term.

2. *Evaluation of antecedents*: for each rule, compute its firing strength by combining the membership values of each term in the antecedent with the conjunction operator.

3. *Implication*: apply the generalized modus ponens to compute the consequent of each fired rule.

4. *Aggregation of consequents*: aggregate the consequents of the fired rules with the union of fuzzy sets.

5. *Defuzzification*: compute the crisp output by applying a defuzzification operator to the aggregated consequent.

For the sake of readability, we skip the formal description of the fuzzy inference process. Detailed analysis of the architecture and of the inference process of Mamdani-type FRBSs can be found in (MA75; Bab02; Ros04).

As stated by Zadeh in (KY96), a Mamdani-type FRBS has two different levels of interpretation.

- A *surface structure*, i.e. the RB, which represents the pure symbolic and logical meaning of the rules.

- A *deep structure*, i.e. the whole KB, which comprises the specific computation-oriented definition of the rules, including their mathematical formulation given by the fuzzy sets associated with the linguistic terms.

Also, we remark that, as stated in (Gui01),

> The strength of fuzzy inference systems relies on their twofold identity. On the one hand, they are able to handle linguistic concepts. On the other hand, they are universal approximators able to perform non linear mappings between inputs and outputs.

Indeed, FRBSs are usually considered as both *interpretable*, i.e., their structure and reasoning process can be easily understood by humans, and *accurate*, i.e., when applied to real-world problems, they achieve performance which are comparable with other state-of-the-art modeling and control techniques, such ad PID controllers, artificial neural networks, Bayesian classifiers, Kalman filters, etc.

### 2.1.4 Hybrid Fuzzy Systems

The most well-known drawback of FRBSs consists in the fact that they are somehow static, i.e., FRBSs are not intrinsically suitable to perform adaptive modeling of real-world data. Hence, recognizing the extremely interesting properties of FRBSs in terms of interpretability, readability and human-oriented representation, various researchers have tried to augment FRBSs with learning and adaptation capabilities.

As stated in Chapter 1, two of the most successful approaches in this field have been the hybridizations achieved in the framework of soft computing, using artificial neural networks and evolutionary algorithms. This hybridization process led to the development of a whole branch of literature in which FRBSs are used as classification, regression and prediction tools built from data. In these solutions, however, the original nice properties of readability and interpretability of FRBSs are usually lost to achieve better accuracy. Indeed, in the past, several researchers have considered FRBSs as black boxes, aiming to obtain perfect models of the training data rather than to extract significant general rules and concepts. This issue, which is extensively addressed in Chapter 4 with a specific focus on CA-related algorithms, is usually referred to as the *accuracy-interpretability trade-off* (CCHM03; CGH$^+$04).

In the last years, this trend has been reversed, and the development of approaches based on interpretability-oriented algorithms has become an extremely popular topic (dO99; Gui01; JGSRB01; CCHM03; Nau03; BS03; CGH$^+$04; PD04; MJG05; WKJ$^+$05; CDLM07; GRP$^+$07; IN07; Ish07).

In the following two Sections, we briefly review *neuro-fuzzy systems* and *genetic fuzzy systems*. For detailed surveys of hybrid fuzzy systems, we refer the interested reader to (Abr01; CGH$^+$04; Her08).

#### 2.1.4.1 Neuro-Fuzzy Systems

Artificial neural networks (ANNs) are computational models that, akin to the structure of human brain, are composed of a network of simple interconnected nodes called *artificial neurons* (Hay94). ANNs have proved to be useful in typical applications performed by the human brain, such as image perception, pattern recognition, regression, classification and prediction. Typically, ANNs are able to learn from data, thanks to learning algorithms which adapt the structure of the network such that the outputs fit the examples given as training set.

As stated above, ANNs are also used in hybridization of fuzzy systems to provide them with learning capabilities. Usually, the hybridization maps the FRBS into an ANN composed of several layers of neurons, where each layer implements a step of the inference process described in Section 2.1.3. Then, standard or ad-hoc learning algorithms are used to modify the parameters of the system so as to model the training data.

The changes induced in the system can simply regard the parameters of the MFs or, in more complex approaches, the overall structure of the system, including the rules and even the implementation of the implication function (Abr01). One of the most popular neuro-fuzzy system is ANFIS (Jan93), which implements a Sugeno-type FRBS, a kind of FRBSs particularly suited for regression applications. Indeed, in a Sugeno-type FRBS, the consequents of the rules are not fuzzy propositions expressed by linguistic terms, as in Mamdani-type FRBSs, but rather $n$-th order polynomial functions. Further, thanks to their architecture, Sugeno-type FRBSs can be seamlessly trained by means of backpropagation-like algorithms.

#### 2.1.4.2 Genetic Fuzzy Systems

Genetic algorithms (GAs) are a gradient-free, general-purpose and population-based meta-heuristic technique employed to solve optimization problems (MS96; HLV98; Mic99; Jon06). GAs evolve solutions to complex problems by imitating natural selection, i.e., the process of adaptation to the environment performed by living beings. Among their most interesting features, we remark the following.

- GAs are a population-based search strategy, i.e., they do not find a look for a single solution to a problem, but rather explore the search space with a set of candidate solutions.

- GAs are able to find “good solutions” to an unconstrained problem in a reasonable time, and they always find at least one “good” sub-optimal solution.

- GAs do not require a differentiable objective function and can be tailored to handle any kind of constraint.

- GSs can solve real-, binary-, or integer-valued problems by choosing a proper representation schema for the chromosome.

- GAs can be adapted in given meta-parameters (e.g., number of individuals in the population), so as to scale well as the problem size increases.

- GAs can be customized to include some heuristics and experts knowledge in the generation of the initial population and in the design of the genetic operators.

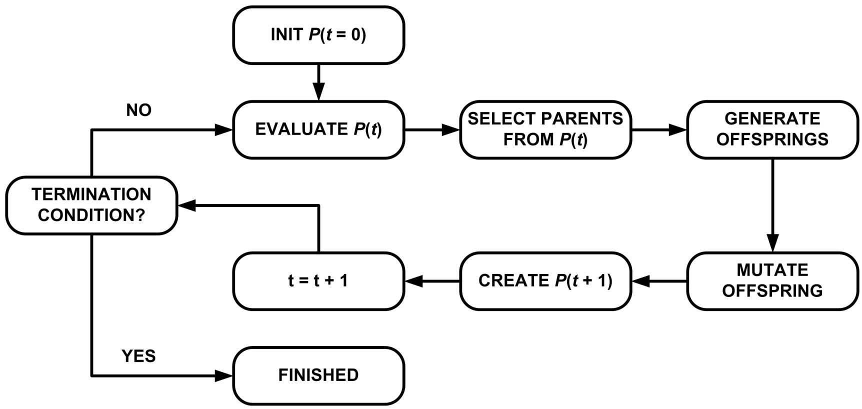



**Figure 3:** The steps of the SGA

As stated above, a GA determines, rather than a single solution, a whole population consisting of *individuals*, which are all candidate solutions to the problem. The distinctive features of each individual are mapped into a structure called *chromosome*. The chromosome is a string of genes, whose values can be chosen in a set of symbols. An application-dependant process generates the individual by decoding its chromosome. Depending on the nature of the problem, the symbols used as values of the genes can be binary, integer or real numbers. Once an individual is generated, a *fitness function* is employed to evaluate its goodness as a solution to the problem. Usually, low fitness values are given to the best individuals (minimization problem). The chromosome is sometimes also called *genotype*, while the decoded version of the solution is referred to as the *phenotype*. For the sake of simplicity, in the following we will blur the definitions of individual and chromosome.

A GA starts at time $t = 0$ with an initial population generated either randomly, or with some heuristic approach that exploits the knowledge of an expert in the problem domain. The algorithm then proceeds in steps called *generations*. At each generation $t$, a new population $P(t + 1)$ is evolved from $P(t)$. As generations pass, the population should improve globally thanks to the application of genetic operators that mimic the natural evolution mechanisms. To this aim, the best individuals are chosen from $P(t)$ (*selection*) to be mated (*crossover*) and slightly modified (*mutation*), so as to create the new population $P(t+1)$.

The selection operator is used to decide which individuals in $P(t)$ should be chosen to generate $P(t+1)$. Optionally, an *elite* of the selected individuals (i.e. a small subset of the best performing individuals) survives and is moved from $P(t)$ to $P(t+1)$ without any change.

The rest of the population is obtained through a crossover operator which chooses some of the individuals and mates them, that is, substitutes them with their offspring, which are newly generated individuals obtained by mixing the genetic material in the parents' chromosomes. The actual implementation of a crossover operation very much depends on the coding schema of the chromosome.

Finally, the mutation operator is invoked to introduce some new genetic material in the population by randomly modifying the values of some genes. Again, different kinds of mutation operators can be defined to handle different sets of symbols.

The population continues to evolve until a stopping criterion is fulfilled, the simplest being a maximum number of generations.

The overall algorithm described above is known as the *simple genetic algorithm* (SGA) and is also depicted in Figure 3.

Note that, while some classes of problems can be solved by directly applying the SGA, more often the development of such an algorithm for a specific problem requires an elaborated engineering process involving a good amount of design and tailoring. Indeed, the typical design of a GA for a given problem includes at least finding suitable representation schemata, coding strategies, genetic operators, and values of parameters.

Recently, sophisticated algorithms based on the paradigm of evolutionary computation have gained popularity for hybridization with FRBSs. For instance, *multi-objective evolutionary algorithms* (MOEAs) like NSGA-II (DAPM02) and PAES (KC00), and *cooperative coevolutionary strategies* (PJ00) have been largely explored. To refer to the set of all evolutionary-inspired techniques such as GAs, MOEAs, cooperative coevolutionary strategies, etc, we will sometimes use the generic term of *evolutionary algorithms* (EAs).

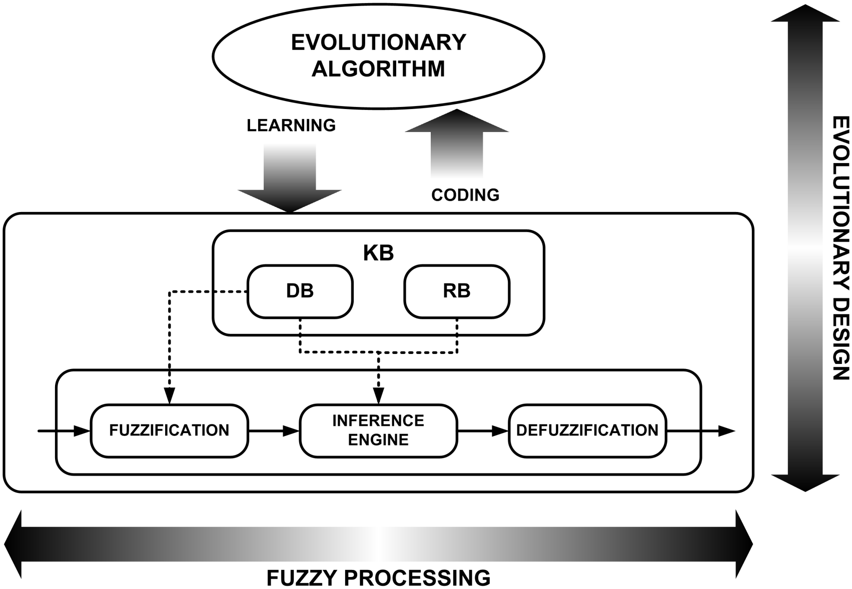



**Figure 4:** The architecture of a GFRBS

A *genetic fuzzy system* (GFS) is a fuzzy system modified by a learning process based on a EA. Similarly, we talk of a *genetic FRBS* (GFRBS) when the hybridized fuzzy system is an FRBS. Figure 4, taken by (CGH$^+$04), shows the typical architecture of a GFRBS.

In a GFRBS, the learning process based on EAs may concern different parts of the FRBS. For instance, some approaches exploit the EA to identify the overall structure of the FRBS (KB and even operators employed in the inference engine), whilst some other just perform a parameter optimization of the MFs comprised in the DB. Obviously, such choice reflects in both the encoding of the individuals (i.e., in the schema of the chromosome) and in the computational complexity of the genetic operators. In (Her08), Herrera proposed a taxonomy of GFRBSs composed by two classes.

- *Genetic tuning* is performed if the RB of the FRBS is previously existing and, therefore, the DB is the only part to be optimized.
- *Genetic learning* is performed if the overall KB has to be identified.

We remark that, as stated by Cordón et al. in (CGH$^+$04), GFRBSs are often preferable to neuro-fuzzy systems. Indeed,

> In some cases, genetic optimization of a fuzzy system is preferable rather than using a neuro-fuzzy approach, because GAs permits a deeper control of the optimization process: to some extent, we could say that, while neuro-fuzzy systems are usually transparent to the designer, in that he/she can neither control the training process nor interpret the final result, GFSs are usually easier to deal with and can be modeled also to cover aspects of user-defined constraints and interaction (including, e.g., interpretability).

GFRBSs have proven to be extremely effective in a number of applications (SR00; RP01; CGH$^+$04; WKJ$^+$05; CDLM07). Hence, several different approaches have been developed in the literature. Whilst early approaches like (Kar91) were based on the SGA described above, in the last years other evolutionary techniques have been applied to GFRBSs so as to handle constrained learning (SR00; CCdJH05), multi-objective optimization (WKJ$^+$05; GRP$^+$07; IN07; Ish07; CDLM07), and even cooperative coevolution (PRS01; DZG04; ZYYLY$^+$06; BDLM08).

As it will be clear in Chapter 4, the shift toward more complex and innovative GFRBSs based on state-of-the-art EAs has mainly been driven by the need for algorithms which could be able to enforce the interpretability constraints usually required in the automatic identification of FRBSs from data.

## 2.2 Context Adaptation of Fuzzy Systems

Context is an extremely influential factor in many areas of science and engineering. A detailed analysis of the possible definition of context was performed by Bazire and Brézillon in (BB05). In their conclusion, they state that

> The context acts like a set of constraints that influence the behavior of a system (a user or a computer) embedded in a given task.

Obviously, such a general definition of context allows for a wide range of research studies in the fields of psychology, linguistic, computer science, and system engineering (Bré02). In this thesis, we focus on the relations between context and system identification, and more precisely on the topic of context adaptation of fuzzy rule based systems. Hence, a detailed analysis of the multi-disciplinary approaches to context that can be found in the literature are out of the scope of the current work. For a survey of existing applications of the notion of context in several areas, we refer the interested reader to the latest proceedings of the *International and Interdisciplinary Conference on Modeling and Using Context* (DKLT05; KRRBV07).

In a sense, *context adaptation* (CA) can be regarded as a knowledge engineering technique that deals with the generation of context-adapted systems from universal models, and vice-versa. A universal model represents versatile knowledge that can be reused in several different environments. On the other hand, a context-adapted system is a specialized version of the model which is properly suited to work in a specific environment.

It can be easily observed that several real-world systems exhibit common behaviors in different environments (PGG97; Tur97). For instance, let us consider the well-known *cart and pole balancing problem* shown in Figure 5. This problem is a common benchmark for control techniques: a cart has an inverted pendulum hinged on its top which need to be balanced by applying lateral forces to the cart. The cart moves over a track,

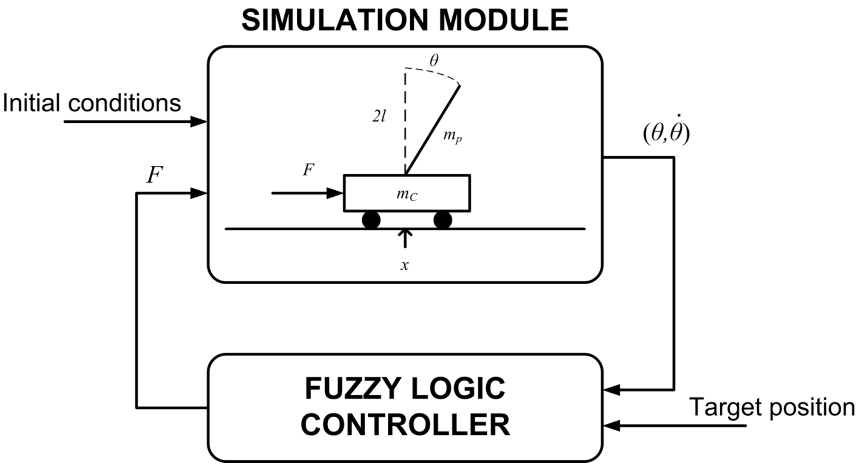



**Figure 5:** The cart-pole system

and thus the pole has one degree of freedom on the vertical plane parallel to the cart (Flo05). A proper controller is needed to automatically perform this task, i.e., to select the amount and direction of the lateral force that is needed to balance the pole. The problem can be made more complex by adding a concurrent second goal, i.e., moving the cart to the center of the track.

The physical behavior of the cart and pole system can be described by

$$\ddot{\theta} = \frac{g \sin\theta + \cos\theta \left( \frac{-F - ml\dot{\theta}^2 \sin\theta}{m_c + m_p} \right)}{l \left( \frac{4}{3} - \frac{m_p \cos^2\theta}{m_c + m_p} \right)}, \tag{2.5}$$

where $\theta$ is the angle between the pole and the vertical, $\mathbf{F}$ is the force applied to the cart, $m_c$ and $m_p$ are the mass of the cart and of the pole, respectively, and $l$ is the half-lenght of the pole.

Although simple, this toy system provides an hard benchmark for learning algorithms due to its strong non linearities. Hence, several approaches based on fuzzy logic controllers have been developed in the literature to solve this problem. It is also remarkable that this system has been one of the first applications of GFRBSs (Kar91).

In the simple version of the problem described by Equation 2.5, the controller usually takes the current values of $\theta$ and $\dot{\theta}$ as inputs and gives the direction and the value of $\mathbf{F}$ as output.

Obviously, the development of a controller for such system must rely on the actual values of the physical parameters, i.e., gravity, length of the pole, and masses. Thus, for each different scenario identified by a specific setup of physical parameters, a different controller is needed or, at least, an existing one has to be properly tuned to respond to the mutated environmental conditions. A different solution may involve the characterization of the four physical parameters as input variables. However, the latter approach seems not convenient for at least two reasons: first, by adding four additional input variables, the problem becomes much more complex than the original two-input one and, second, physical parameters usually change much less frequently than other variables, and, therefore, should be approached differently.

Actually, the physical parameters of the system represent the external context where the FLC should operate. From the point of view of a human controller, there is no need to know exactly the values of such parameters: indeed, the rules exploited by a human controller are mostly derived from experience and are independent of the external context. Further, such rules can be expressed, as it is common in FRBSs, via linguistic terms. For instance, typical rules that might be exploited in a cart-pole FLCs are

$$\begin{aligned}&\text{if } \dot{\theta} \text{ is } \textit{strongly positive} \text{ and } \theta \text{ is } \textit{strongly positive}\\&\text{then } \mathbf{F} \text{ is } \textit{strongly positive},\end{aligned} \tag{2.6}$$

and

$$\begin{aligned}&\text{if } \dot{\theta} \text{ is } \textit{small positive} \text{ and } \theta \text{ is } \textit{strongly negative}\\&\text{then } \mathbf{F} \text{ is } \textit{negative}.\end{aligned} \tag{2.7}$$

Obviously, a human controller will adapt the force applied to the cart with respect to the external context, but still following the intuitive rules that are universal and completely context-independent. Therefore, the external context affects just the meaning associated with the degrees of intensity (e.g., *strongly positive* in the example of Equation 2.6) associated with the force.

Depending on the application, context may be determined by several factors, including the external environment, the capabilities or the features of the observer/controller, the interaction with other systems, etc.

Another significant example of how context affects the behavior of a system can be found in (PGG97). Let us consider how humans learn to drive a car. Once one has learnt to drive a given model of car in a given weather condition, he/she is able to drive various other models of cars in different weather and traffic conditions, without having to learn again from scratch all the driving rules. Rather, he/she will just have to adapt his/her driving style to the new context determined by the environment.

As an example, suppose one has learnt to drive in normal weather conditions and at an average speed of 30 mph. Hence, when the weather is rainy, he/she will use the same set of rules (e.g., crisp rules such as *if light is red then stop* and fuzzy rules such as *if speed is high then pressure on throttle is small*), but will adapt the range of speed around the average value of, for instance, 20 mph. Similar adaptations may occur when driving during nighttime or when changing model of car.

Other examples of CA could be found in popular applications of fuzzy systems in embedded systems, such as auto-focus devices in cameras, anti-lock braking systems and cruise controllers in cars, etc.

Hence, the concept of context is critical to assess the *reuse* of already existing controllers or models, thus leading to a speed-up in the development of new systems when a different version of the same system is already available for another context. Indeed, learning the RB is usually the most expensive process of FRBS automatic identification. To sum up, as stated by Pedrycz et at. in (PGG97),

> the key point [...] is that the rules [...] are expected to be universal to a high extent.

It is remarkable to notice that context can be actually expressed in several different ways. For instance, Brézillon in (Bré02) defines it in terms of *patterns*, as

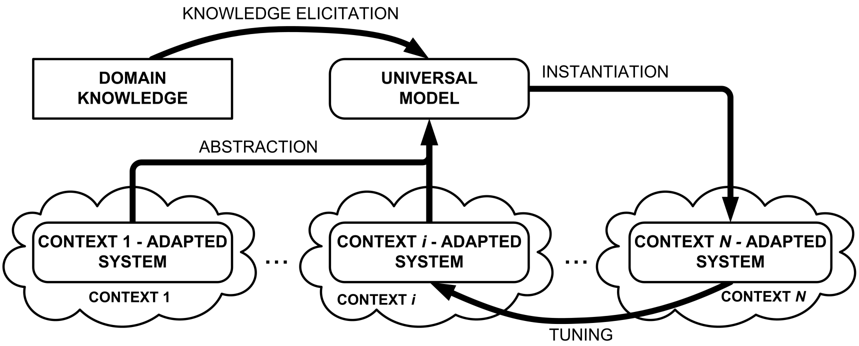



**Figure 6:** The conceptual framework of CA

> Each model describes a problem that recurs the environment and thus describes the heart of the solution to this problem in a way that permits the reuse of the solution without to do it twice exactly in the same way [...] A pattern is a proven solution to a problem in a context.

On the other hand, Gudwin et al. in (GGP98) address context from the observer's point-of-view, relating the effect of context to the perception of *stimuli*.

> The same stimulus information in different contexts can produce different perceptual events [...] The main effect of a context can be related with some sort of filtering. That is, the same base concept can be perceived in different situations, provided it is filtered to suit the context particularities.

Finally, another interpretation, introduced by Magdalena in (Mag02), relates context to *actions*.

> (These) actions [...] in some sense are generic actions. The context [...] will translate those generic actions into specific actions adapted to the situation.

In the framework of system identification, the idea that stands behind context adaptation is the extraction of universal models from each system or from domain knowledge, so as to use them, in adapted form, when facing a similar problem in another environment.

The overall conceptual framework of CA is depicted in Figure 6. Here, the universal model and the context-adapted systems are represented by rounded rectangles, whereas the knowledge related to the application domain is depicted by a box. The four arrows represent the following processes.

1. *Abstraction*: an application-dependent process, typically data-driven, that is used to extract a universal model from given systems.
2. *Knowledge elicitation*: a heuristic process in which domain experts explicitly translate their expertise into a universal model for a given pattern of problems.
3. *Instantiation*: the inverse process of abstraction, in which a universal knowledge is augmented with context-dependent knowledge in order to suit it to the target environment.
4. *Tuning*: the simple reuse of existing knowledge from a context to another one, without employing a universal model.

The four processes listed above are employed to generate context-adapted systems and universal models. Obviously, the actual techniques and algorithms used to implement the processes are strongly application-dependent.

FST, FL, and fuzzy systems provide a fertile background for the application of the proposed CA framework. The knowledge elicitation and the instantiation processes depicted in Figure 6 have been successfully applied to the development of fuzzy systems and, more specifically, to Mamdani-type FBRSs. Indeed, in Mamdani-type FRBS identification, the RB is often derived from heuristic knowledge, which is usually valid independently of the real environment where the FRBS will work. Hence, as we will detail in the following Sections, the RB can be considered as a context-free model. In other words, the real environment does not affect the RB (i.e., the surface structure of the FRBS), but rather influences the DB, and, more specifically, the meaning associated with each linguistic term used in the rules.

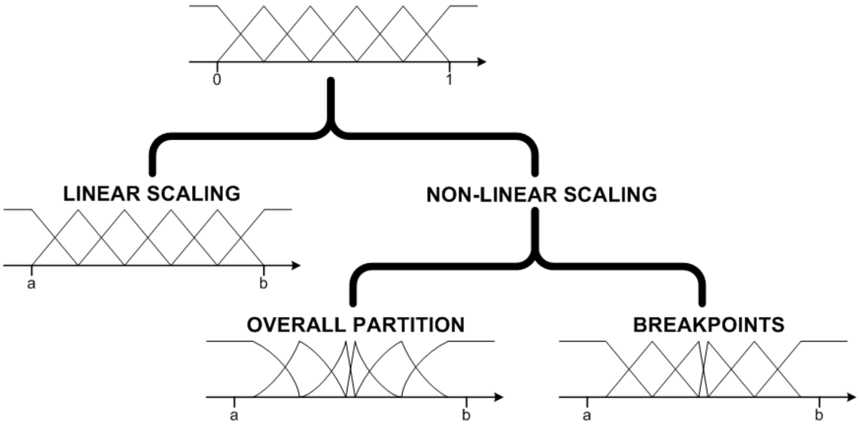


**Figure 7:** Examples of application of different types of scaling functions to a normalized partition

### 2.2.1 Previous Work

In the literature, the majority of papers on CA of FRBSs have focused on the use of *scaling functions* (Bas94; GG94; Mag97; PGG97; GGP98; CHMV01; Mag02; CdJH$^+$03). A scaling function serves to adapt a fuzzy partition by mapping an existing universe of discourse to the context-adapted universe, possibly modifying the distribution and the shape of fuzzy sets. Usually, the scaling function is applied to a normalized partition, i.e., a partition defined over the $\mathcal{N} = [0, 1]$ universe of discourse and uniformly partitioned into triangular, trapezoidal or Gaussian fuzzy sets. The number of fuzzy sets for each partition coincides with the number of linguistic terms defined for the linguistic variable corresponding to the universe of the partition.

The scaling functions used in the literature can be classified into *linear* (Bas94; GG94; Mag02) and *non linear* (Mag97; PGG97; GGP98; CHMV01; CdJH$^+$03; Kla06). Non linear scaling functions can be applied on the overall universe of discourse, thus modifying the shape of fuzzy sets, as in (Mag97; PGG97; GGP98; Kla06), or just on some points (e.g., on breakpoints in the case of triangular and trapezoidal fuzzy sets), as in

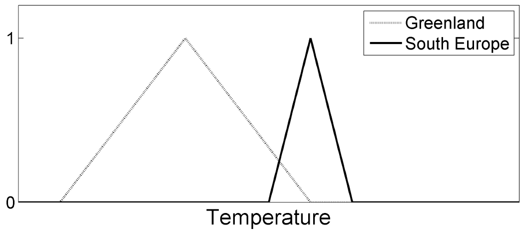



**Figure 8:** Two possible meanings of the linguistic term *hot*: in Greenland (dotted) and in South Europe (solid)

(CHMV01; CdJH$^{+}$03), so as to maintain the original shape of the fuzzy sets and the interpretability of the partition. The different possible types and applications of scaling functions are summed up in Figure 7.

The first approaches to CA of fuzzy systems date back to mid 1990's (Bas94; GG94). In (GG94), Gudwin and Gomide stated that the concept of context mainly relies on the idea of restriction. Indeed, the terms of a linguistic variable are intrinsically relative if not instantiated to a specific context, i.e., if the universe of discourse $\mathcal{U}$ of the variable is not clearly defined. The actual binding between terms and context is given by the restriction of the working range of the universe of discourse. In other words, by setting the actual bounds of the universe of discourse of a linguistic variable, we also fix the parameters of its fuzzy sets and, thus, we define the exact meaning of the linguistic terms. This operation can be performed by a linear mapping from a context-free universe of discourse to the context-adapted one. In the former universe, terms are defined only by implicit linguistic relations among them.

For instance, the linguistic term *hot* can have different meanings (i.e., different associated fuzzy sets), depending on the geographical and cultural context. Obviously, an Eskimo and a South European would perceive the concept of heat in different ways. However, both of them will agree that the linguistic term *hot* is somehow "greater" than the linguistic term *cold*, independently by the context. Figure 8 shows two possible instances of the term *hot* at different latitudes.

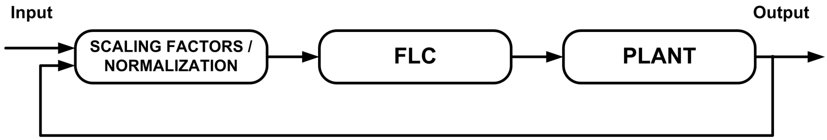



**Figure 9:** An FLC in a simple feedback control system

Gudwin and Gomide proposed some techniques to determine the working range of a linguistic variable in a specific context from real-world data, namely the *absolute limit context determination*, the *elastic limit context determination*, the *statistic context determination* and a *neural-network-based context generation*. Each of the proposed techniques can be used for an off-line adaptation (in the case all the data are available simultaneously) and for an on-line adaptation (in the case data are temporally distributed).

We can represent the context-free meaning of linguistic terms by fuzzy sets defined in the normalized $\mathcal{N} = [0, 1]$ universe of discourse. After the identification of the actual bounds $\mathcal{U} = [u_{\min}, u_{\max}]$ of the linguistic variable, the fuzzy sets are fixed in the context-adapted universe of discourse by means of a linear scaling function

$$\varphi_0(x, u_{\min}, u_{\max}) : [0, 1] \to \mathcal{U} = u_{\min} + (u_{\max} - u_{\min})x. \tag{2.8}$$

A similar approach based on linear scaling of linguistic variable was used by Bastian in (Bas94) to heuristically develop an ad-hoc controller for automatic gear selection on cars, where the different contexts are defined by the drivers style (sporty or average). Bastian made use of *flexible linguistic variables*, i.e., linguistic variables defined over a universe of discourse without predefined bounds.

Determining the bounds of a linguistic variable is topic widely explored also in fields different from context adaptation, such as FLC identification. A typical use of an FLC in a simple feedback control system, taken from (Ros04), is shown in Figure 9.

Since an FLC is usually developed to work on a fixed range of inputs and outputs (either in the $[0, 1]$ or in the $[-1, 1]$ universes of discourse), inputs coming from the plant and/or from the external environment need to be normalized in order to let the FLC compute the output. The multipliers that are used to perform the normalization are called *scaling factors* (WCL00; CHMV01; Ros04; DW05). Even though there is no explicit concept of CA in this process, determining the values of the scaling factors for a specific physical scenario is equivalent to instantiating the actual universe of discourse of the input and output variables (DW05) and so to context-adapting the FLC. Remarkable contributions on this topic can be found in (MP99; WCL00; DW05), where self-tuning approaches to the identification of scaling factors are proposed.

In (Mag97), Magdalena introduced a non linear scaling function to instantiate an FRBS to a specific context. Unlike linear scaling functions, non linear scaling functions not only fix the parameters of the fuzzy sets so as to change the working range of the universe of discourse, but also modify their shape and distribution in the space.

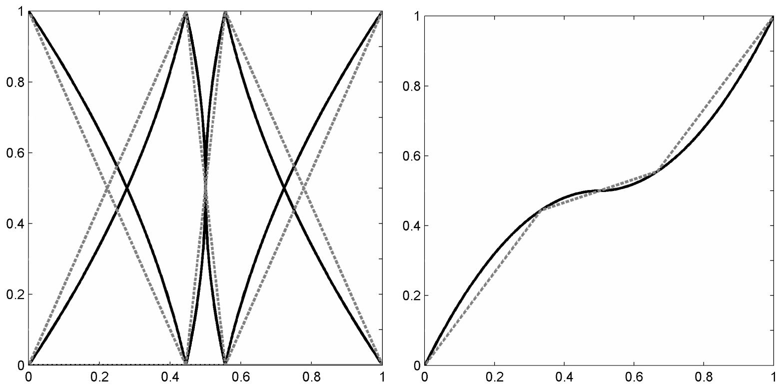
Unfortunately, an infinite number of non linear scaling functions may be used to perform such adaptation. To achieve a more efficient exploration of the search space, Magdalena proposed to restrict the candidates to set of functions belonging to the family identified by

$$\varphi_1(x, u_{\min}, u_{\max}, a) : \mathcal{U} \to [-1, 1] = \mathrm{sign}(x')|x'|^a, \tag{2.9}$$

where $x' = 2\frac{x-u_{\min}}{u_{\max}-u_{\min}} - 1$, and $a > 0$ is a *sensitivity parameter* which determines the intensity of the concentration ($a < 1$) or of the dilation ($a > 1$) of the fuzzy sets. The parametrized function $\varphi_1$ first performs a linear mapping from the context-adapted universe $\mathcal{U}$ to the normalized universe $\mathcal{N}$, and then introduces the non linear distribution by concentrating or dilating the fuzzy sets around $0$.

Note that a scaling function can either be defined as $\mathcal{N} \to \mathcal{U}$ (like $\varphi_0$), if it acts on the normalized universe of discourse, or as $\mathcal{U} \to \mathcal{N}$ (like $\varphi_1$), if it acts on the universe of discourse of inputs and outputs.

To instantiate a context-adapted Mamdani-type FRBS, Magdalena exploited a genetic learning process (CGH$^{+}$04) that, aiming to minimize the error with respect to real-world data, is able to concurrently find an optimal set of rules and the optimal values of the parameters of the scaling function for each input and output linguistic variable.




(a) Linguistic variable (b) Scaling function $\varphi_2$

**Figure 10:** Application of $\varphi_2$ with $a = 0.5$ to a fuzzy partition: to the whole universe (solid) and only to breakpoints of the fuzzy sets (dashed)

An augmented version of $\varphi_1$, denoted $\varphi_2$ in the following, was introduced in (CHMV01; CdJH+03). Function $\varphi_2 : \mathcal{U} \rightarrow [-1, 1]$ allows the concentration and dilation of fuzzy sets not only around the center of the normalized universe of discourse, but also around the extremes $-1$ and $1$. The function, which is described by a four-step process, takes two parameter as input: the binary parameter $S \in \{0, 1\}$, used to determine the chosen point for concentration or dilation, and the real-valued parameter $a > 0$, which has the same role as in $\varphi_1$.

In (CHMV01), $\varphi_2$ was not applied to each input and output value of the system, but rather just to the breakpoints of the triangular fuzzy sets. Thus, the distortion introduced in the universe of discourse changes the distribution of the fuzzy sets, but not their shape. This is equivalent to applying a piece-wise linear scaling function, which changes slope in correspondence of the breakpoints of the fuzzy sets. Figure 10 shows a comparison of the application of $\varphi_2$ with sensitivity $a = 0.5$ to the whole universe and only to the breakpoints of fuzzy sets.

The concepts of CA of linguistic variables introduced in (GG94) has been later expanded by Pedrycz et al. in (PGG97; GGP98). In (PGG97), the authors developed two relevant contributions to CA of FRBSs.

First, in an FRBS, the RB should be considered as universal knowledge, independently of the specific context, and thus should be reused in a number of similar situations. Indeed, rules express a logical relation between terms, whose meaning depends on the context, but whose validity is absolute. As an example, let us consider the following rule

$$\text{if } \textit{temperature} \text{ is } \textit{hot} \text{ then } \textit{fan speed} \text{ is } \textit{high}.$$

Although the meaning of the terms *hot* and *high* is strongly context-dependent, the rule has a universal meaning because it expresses a logical relation between terms of linguistic variables. Also, the terms show an implicit ordering that is context-free. For instance, as already stated above, the term *cold* must always precede the term *hot*. Therefore, the knowledge coded in the RB should be carefully chosen so that it can be considered universal and, thus, should not be modified by, e.g., the instantiation process. It follows that CA of FRBSs should concentrate on linguistic variables and on fine tuning of the fuzzy sets so as to optimally fit a specific context. To this aim, the authors initially considered a linguistic variable in the normalized $\mathcal{N} = [0, 1]$ universe of discourse, where linguistic terms are instantiated to uniformly distributed fuzzy sets. Like in (Mag97), the mapping from the context-free universe of discourse to the context-adapted one was performed by means of non linear scaling functions $\varphi$, that act as $\varphi : \mathcal{U} \to \mathcal{N}$.

Second, the authors introduced a set of requirements for the definition of proper scaling functions, so as to preserve ordering and normality of the original linguistic variable. In particular, they required *continuity*, *non-decreasing monotonicity*, and the *boundary conditions* $\varphi(0) = u_{\min}$ and $\varphi(1) = u_{\max}$. Furthermore, *differentiability* is required when a learning algorithm is used to determine the optimal parameters of the scaling function.

In (PGG97), the scaling function was implemented by a two-layered feed-forward neural network trained with contextualized real-world data. In (GGP98), a genetic learning process is used to find the optimal parameters of scaling function $\varphi_3$, defined as a linear combination of $c$ sigmoidal functions

$$\varphi_3(x, \mathbf{k}, \mathbf{m}, \mathbf{s}) : \mathcal{U} \to [0, 1] = \alpha \sum_{i=1}^{c} \frac{k_i}{1 + \mathrm{e}^{\frac{-(x-m_i)}{s_i}}} + \beta, \tag{2.10}$$

where $\mathbf{k} = (k_1, ..., k_c)$, $\mathbf{m} = (m_1, ..., m_c)$, $\mathbf{s} = (s_1, ..., s_c)$, and $k_i \geq 0$, $s_i > 0$, and $m_i \in \mathcal{U}\ \forall i = 1, ..., c$. Values $\alpha$ and $\beta$ are chosen so as to respect the boundary conditions defined above.

Some of the CA techniques introduced in (GG94; PGG97; GGP98) were applied to the development of a fuzzy controller for an elevator group in (GGN98).

As regards scaling functions, Klawonn (Kla06) proposed a hybrid learning algorithm to adapt zero-order Sugeno-type FRBSs (Ros04) using a scaling function defined as

$$\varphi_4(x, a, b) : \mathbb{R} \to \mathbb{R} = \frac{1}{1 + \mathrm{e}^{-b(x-a)}}. \tag{2.11}$$

The learning algorithm introduced by Klawonn can concurrently optimize the parameters $a$ and $b$ of $\varphi_4$ and the consequents of the rules in the FRBS. Klawonn remarked that the use of scaling functions can help to reduce the number of parameters of the FRBS that have to be identified in a learning algorithm. Indeed, scaling functions act on the overall universe of discourse, rather than on single parameters of the fuzzy sets. Thus, differently from other fuzzy system tuning techniques, such as, for instance, (Gui01; CCHM03), the number of parameters that have to be optimized scales linearly with the number of input and output variables.

Besides the formulation of CA as scaling-function-based tuning of linguistic variables, other approaches have been developed in the literature. A remarkable contribution by Turner (Tur97; Tur98) introduced concepts related to CA of fuzzy sets into the more general approach of *context-mediated behavior* (CMB). CMB is an approach to the control of intelligent agents that heavily relies on context-related knowledge. Turner used CMB to develop a controller for an autonomous underwater vehicle, named Orca. In CMB, context is presented to the agent via *c-schemas*, that are structures which encode the knowledge needed to solve a specific class of problems. The knowledge is coded into rules, and fuzzy sets

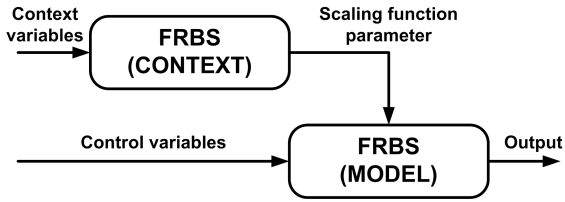



**Figure 11:** The hierarchical system proposed in (Mag02)

are used to represent linguistic terms. Again, in Turner's approach the rules are considered to code a behavior that is general and context-free, while the specific meaning of a term is context-dependent and can be defined by setting the parameters of the associated fuzzy set. For instance, the term *nominal* of linguistic variable *Depth* can take different values in a harbor or in open ocean. C-schemas provide both a way to characterize each context, that is initially described by some expert's static knowledge, and to derive new contexts from the aggregation of existing ones.

Another approach to the dynamic generation of context was developed by Magdalena in (Mag02). Starting from a simple linear scaling function adaptation, the author proposed the use of a two-level hierarchical FRBS model (Figure 11), where the first level defines scaling factors used in the universes of the input variables of the second level. A related technique can be found in (DW05). In a sense, the FRBS in the first level represents a model of the context itself: indeed, it is able to interpolate the parameters of unknown contexts by evaluating a subset of the input variables that strongly influence the context.

Finally, it is remarkable to note that a different notion of context, not strictly related to CA of FRBSs, has also been explored in FST literature for the definition of context-aware linguistic hedges (CK02; HN04).

### 2.2.2 A Common Framework

As we have described in Section 2.2.1, a standard approach to CA of FRBSs has emerged from previous work.

1. The RB is considered as a context-free and universal knowledge.

2. The linguistic variables in the DB are context-adapted (i.e., instantiated) by means of some parameterized operators (scaling functions and/or fuzzy modifiers) that act on the overall universe of discourse.

To determine the optimal configuration of the parameters used in the scaling functions, evolutionary algorithms have been extensively applied (Mag97; GGP98; CHMV01). Most of the earliest approaches rely on heuristic methods (Bas94; GG94; Tur97; Tur98; Mag02), whereas few works explored the use of other soft computing techniques, such as neural-network-like learning (PGG97; Kla06).

Based on the elements and processes comprised in the general framework of Figure 6, we can summarize CA of FRBSs as follows.

1. The *universal model* can be represented by an RB (i.e., by the surface structure of an FRBS), or by a deeply structured FRBS, where fuzzy sets are uniformly distributed and normalized in the context-free $\mathcal{N} = [0, 1]$ universe of discourse.

2. The *context-adapted systems* can be represented by deeply structured FRBSs, where linguistic variables are tuned with a global approach so as to coherently model the effects of context and to reduce the search space.

3. The *knowledge elicitation process* can be heuristically performed by asking domain experts to select general rules to be included in the RB of the universal model.

4. The *abstraction process* can be performed by applying a learn-by-example algorithm, such as the popular one introduced in (WM92), to a (set of) context-adapted systems.

5. The *instantiation process*, on which most of the previous work has concentrated, can be performed by a learning algorithm that finds a proper configuration of parameters of adaptation operators.

6. The *tuning process* can be considered as a special case of instantiation, in which the initial surface structure is not derived from the universal model but rather from another context-adapted system.

The above-mentioned assumption about universality of the RB trivially implies that the RB should not be modified during the adaptation process. Besides this requirement, the semantic relation existing between the RB and the DB forces more constraints on the instantiation of the DB.

First, since each linguistic term is bound to a fuzzy set, each fuzzy set is useful and meaningful only if the associated linguistic term is used in the RB. Thus, the number of fuzzy sets in each partition is directly determined by the number of linguistic labels defined in the RB and, therefore, should not change during CA.

Second, when experts define the RB, they use an implicit semantic ordering of linguistic terms that is significant to humans. For instance, the linguistic term *high* will always follow *low*, and *cold* will always precede *hot*. The ordering of linguistic terms is usually modeled in fuzzy partitions by some ordering of fuzzy sets. Consequently, to preserve semantics, the post-CA ordering of fuzzy sets should reflect the pre-CA ordering of linguistic terms.

To summarize, each CA approach should comply with the following guidelines.

1. CA should not modify the RB.

2. CA should not change the number of linguistic terms defined in the RB and, consequently, the number of corresponding fuzzy sets.

3. CA should not affect the semantic ordering of linguistic terms.

As we will show in the following Chapters, scaling-function-based approaches do not provide enough modeling capabilities to achieve an effective CA. Therefore, in Chapter 3 we introduce a novel CA technique based on improved adaptation operators. By augmenting the modeling capabilities, the instantiation of the fuzzy sets is performed with a higher degree of freedom. On the other hand, such augmented flexibility may generate uneasily interpretable linguistic variables. Hence, in Chapter 4 we discuss how interpretability of an FRBS can be affected by CA and how this issue can be overcome.

We remark that, among the four processes described in Figure 6, only the instantiation process was substantially addressed by previous work. Since this thesis aims to develop a common framework for existing approaches and to expand the original definition of CA of FRBSs, in the following Chapters we will restrict our analysis to the specific process of instantiation. Therefore, we will sometimes blur the difference between the overall framework and instantiation and, in line with the literature, we may cite CA of FRBS to refer to the instantiation process. However, in Chapter 7, we will discuss future work which we intend to carry on so as to develop algorithms to automate the abstraction process.

# Chapter 3

# Novel Operators for Context Adaptation of Fuzzy Partitions

In this Chapter, we introduce five novel tuning operators used to perform the instantiation of a universal FRBS to a specific context. The first operator is a non linear scaling function which performs an adjustment of the universes, so as to cover all possible input and output values and, possibly, to make granularity finer in some parts of the universe and coarser elsewhere. As explained in Chapter 2.2, this adjustment is necessary because CA starts from a normalized FRBS, where the universe of each variable is defined in $[0, 1]$, and uniformly partitioned with trapezoidal MFs.

The other four operators belong to the family of fuzzy modifiers (see Definition 21): they modify the core, the support, and the shapes of each MF. The modifiers are chosen and formulated in such a way that the effects on the resulting fuzzy sets are independent of the order in which modifiers are applied. Thus, modifiers can be applied without interfering with each other. As we will detail in Chapter 5, in our adaptation process we first apply the scaling function, and then the fuzzy modifiers.

Our modifiers are enough flexible to be used in a wide range of tuning applications, not necessarily related to CA, and might be applied to a single fuzzy set. However, in line with previous work, we apply each modifier with the same intensity to the overall partition. This allows us to use a very small number of parameters to represent a wide range of configurations through the combined effects of the five operators.

The fuzzy modifiers are defined so as to act on a trapezoidal fuzzy set $A$ defined by $\mathcal{C}(A) = [cl, cu]$ and $\mathcal{S}(A) = [sl, su]$ on $\mathcal{U} = [u_{\min}, u_{\max}]$, where $sl$ and $su$, and $cl$ and $cu$ are the left and right bounds of the support and of the core, respectively, with $sl \leq cl \leq cu \leq su$. A trapezoidal fuzzy set $A$ is characterized by the following MF

$$A(x; sl, cl, cu, su) = \max\left(\min\left(\frac{x - sl}{cl - sl}, 1, \frac{su - x}{su - cu}\right), 0\right). \tag{3.1}$$

This definition includes the special case of triangular fuzzy sets, i.e. $cl = cu$. Further, we admit the special case $sl = cl = cu = su$, in which the trapezoidal MF degenerates into a singleton, i.e. an MF such that

$$A(x; sl, cl, cu, su) \begin{cases} 1 & \text{if } x = sl = cl = cu = su \\ 0 & \text{elsewhere.} \end{cases} \tag{3.2}$$

Although we use trapezoidal fuzzy sets, we remark that the proposed modifiers can be easily adapted to work on any other shape of fuzzy sets.

## 3.1 Scaling Function

The covering of the universe can be obtained by using any scaling function that satisfies the continuity, monotonicity and boundary requirements suggested in (PGG97). As stated in Section 2.2.1, the first scaling functions proposed in the literature simply performed a linear mapping from the normalized initial universe to the real universe, as in (GG94; Bas94). Though this approach maintains the interpretability of the normalized partition, its modeling capabilities are quite limited.

In (Mag97; CHMV01), the authors propose non linear scaling functions which can concentrate/dilate the MFs around a single point, which can be either the center of the universe or one of the two extremes. Since the choice is limited to only three points, the adaptation capability of these scaling function is restricted to few cases.

In (Kla06) and (GGP98), non linear scaling functions are implemented by a sigmoidal function and a linear composition of sigmoidal-like functions, respectively. While the first approach does not guarantee to preserve normality of fuzzy sets, the second requires several parameters, whose identification can be a long and difficult task.

To overcome these limits, we introduce a scaling function which, extending the approach proposed in (CHMV01), can both map the universe of discourse from a normalized interval $\mathcal{N} = [0, 1]$ to any context-specific interval $\mathcal{U} = [u_{\min}, u_{\max}]$ and non uniformly distribute the MFs, allowing to select any point in $\mathcal{U}$ as center of gravity.

**Definition 36 (Non linear scaling function)** *The non linear scaling function $\psi$ is defined as*

$$\psi(x) : \mathcal{N} \to \mathcal{U} = \begin{cases} u_{\min} + (u_{\max} - u_{\min})(\lambda^{1-k_{SF}} x^{k_{SF}}) & \text{if } x \leq \lambda \\ u_{\min} + (u_{\max} - u_{\min})[1 - (1-\lambda)^{1-k_{SF}} (1-x)^{k_{SF}}] & \text{if } x > \lambda, \end{cases} \tag{3.3}$$

*with $\lambda \in [0, 1]$ and $k_{SF} > 0$.*

Parameter $\lambda$ identifies the center of gravity in the normalized partition, whilst parameter $k_{SF}$ defines the degree of dilation ($k_{SF} > 1$) or compression ($k_{SF} < 1$) of fuzzy sets around $\lambda$.

We observe that, when $\lambda = 0$, $\lambda = 0.5$ and $\lambda = 1$, the adjustments carried out by Equation 3.3 are similar to those presented in (CHMV01). To maintain the original trapezoidal shape of MFs, we do not apply the scaling function to all points of the universe, but rather only to the breakpoints of the MFs, as in Figure 10.

An as example, let us consider the partition $P_{\mathcal{N}} = \{A_1, ..., A_5\}$, shown in Figure 12, composed of $N = 5$ uniformly distributed trapezoidal fuzzy sets on $\mathcal{N}$. Each fuzzy set, except for the first and the last, is characterized by a width of the core and of the support equal to $\frac{1}{2(N-1)}$ and $\frac{3}{2(N-1)}$, respectively. The first and the last fuzzy sets are characterized by a width of the core and of the support equal to $\frac{1}{4(N-1)}$ and $\frac{3}{4(N-1)}$, respectively.

We applied the non linear scaling function $\psi(x)$ with $\mathcal{U} = [0, 1]$ and with different values of $\lambda$ and $k_{SF}$ to $P_{\mathcal{N}}$, namely $\lambda = 0.5$ and $k_{SF} = 2$,

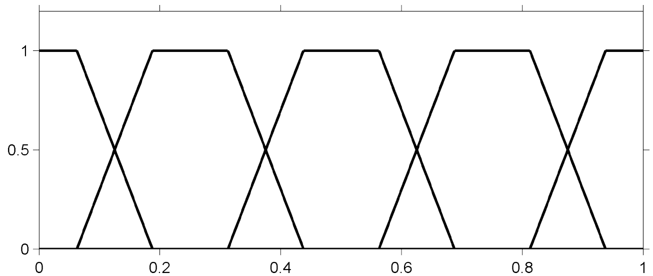



**Figure 12:** The normalized fuzzy partition $P_{\mathcal{N}}$ composed of five uniformly distributed trapezoidal MFs

$\lambda = 0.25$ and $k_{SF} = 0.3$, $\lambda = 1$ and $k_{SF} = 1.5$, and $\lambda = 0.5$ and $k_{SF} = 0.5$. Figures 13 – 16 show the shape of $\psi(x)$ with the chosen values and the effects of its application on the partition.

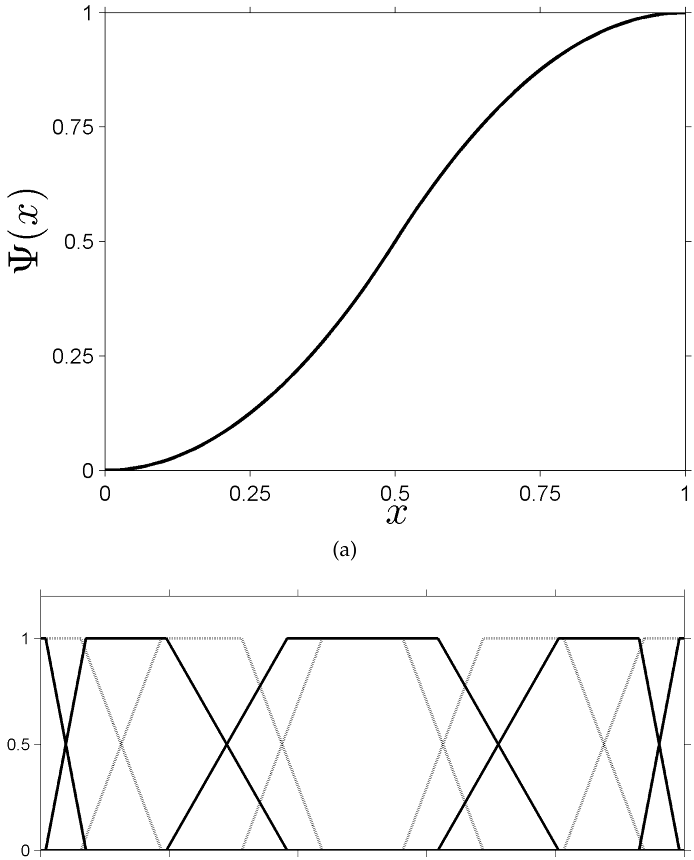
Figures 17 and 18 show how the behavior of the non linear scaling function changes with different values of $k_{SF}$ and $\lambda$. More precisely, Figure 17 depicts the effects of varying $k_{SF}$ in $\{0.25, 0.75, 1.5, 3\}$ for $\lambda = 0.5$ and $\lambda = 1$, whilst Figure 18 depicts the effects of varying $\lambda$ in $\{0.25, 0.5, 0.75, 1\}$ for $k_{SF} = 0.5$ and $k_{SF} = 2$.

## 3.2 Fuzzy Modifiers

In this Section, we introduce the four novel fuzzy modifiers that we designed to perform the instantiation of a universal FRBS to a specific context. To define the fuzzy modifiers, we analyzed the possible variations undergone by the MFs in order to adapt themselves to a context.

We realized that these variations could be reproduced by applying four simple operations: core shifting, core expanding/shrinking, support expanding/shrinking and shape modifying. Thus, we defined four different fuzzy modifiers so as to reproduce these simple operations.

Though the definition of fuzzy modifiers as any mapping $\mathcal{F}(\mathcal{U}) \rightarrow \mathcal{F}(\mathcal{U})$ allows for the definition of very powerful modifiers, we tried to introduce mappings that model the effects of the context without affecting the interpretability of the final partitions and the ordering of the linguistic values.

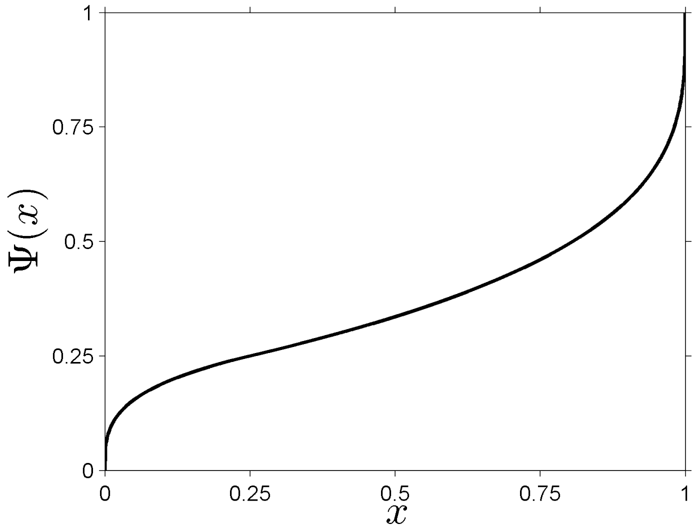



(a)

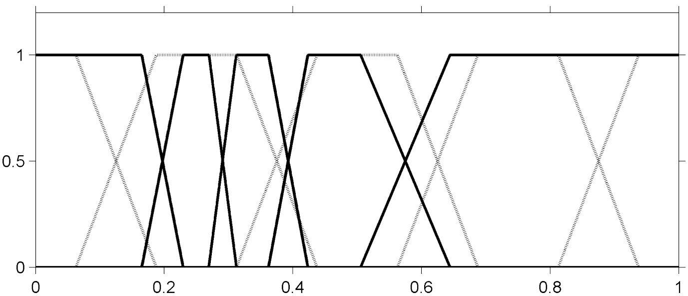


(b)

**Figure 13:** The non linear scaling function $\psi$ and its application to $P_{\mathcal{N}}$ with $\lambda = 0.5$ and $k_{SF} = 2$

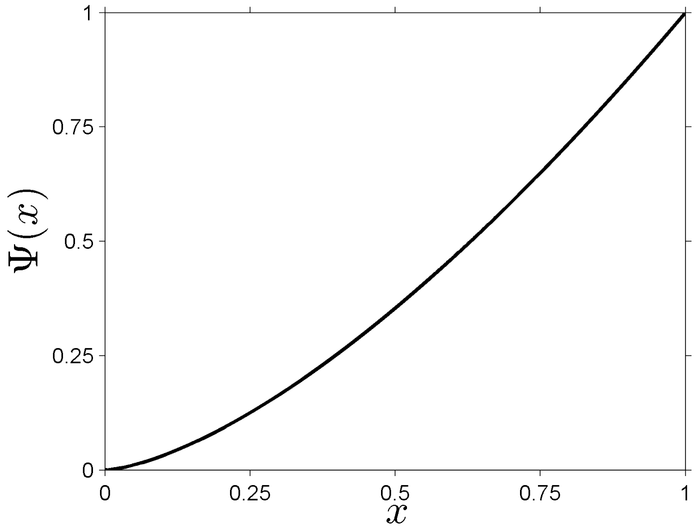



(a)

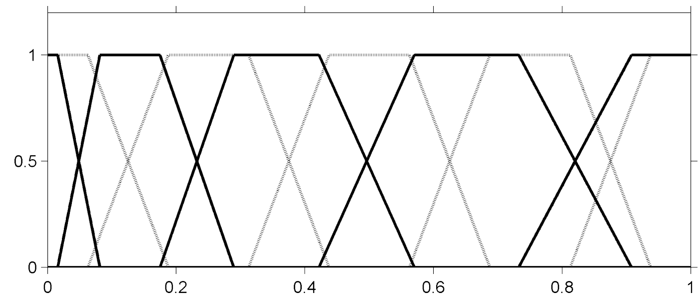



(b)

**Figure 14:** The non linear scaling function $\psi$ and its application to $P_{\mathcal{N}}$ with $\lambda = 0.25$ and $k_{SF} = 0.3$

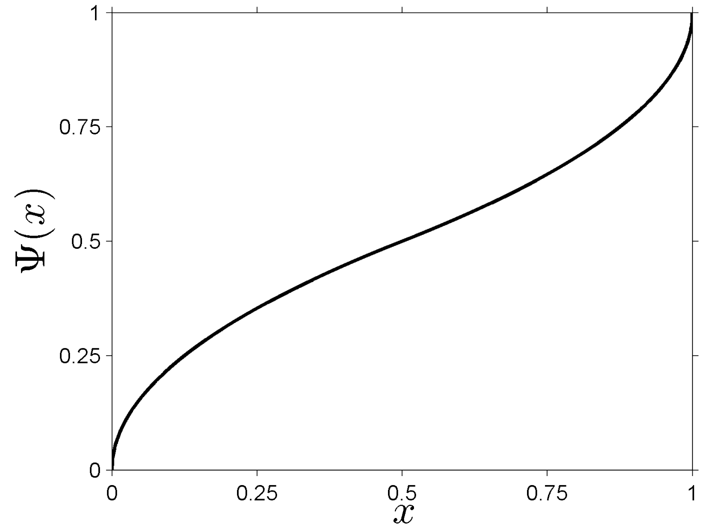



(a)

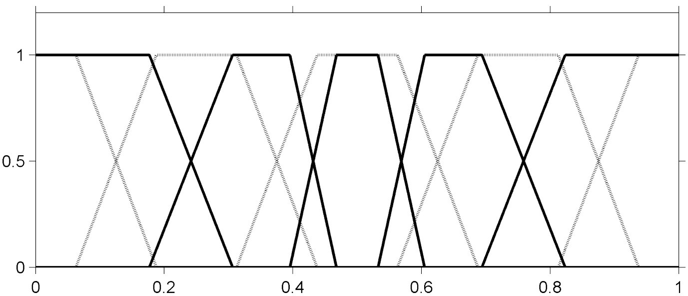



(b)

**Figure 15:** The non linear scaling function $\psi$ and its application to $P_{\mathcal{N}}$ with $\lambda = 1$ and $k_{SF} = 1.5$

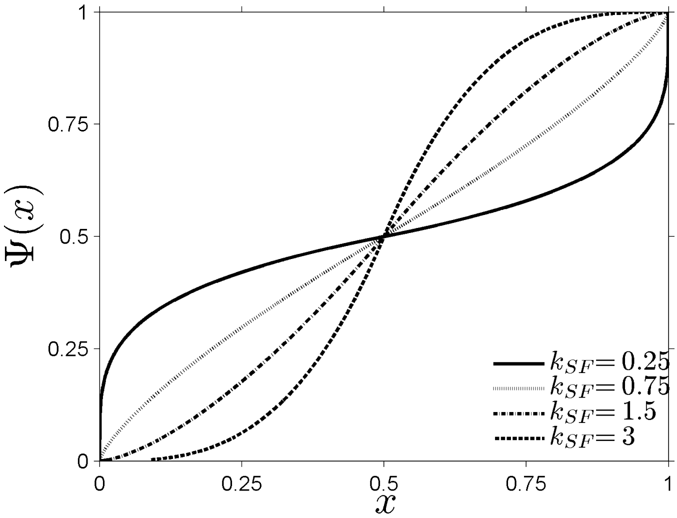



(a)

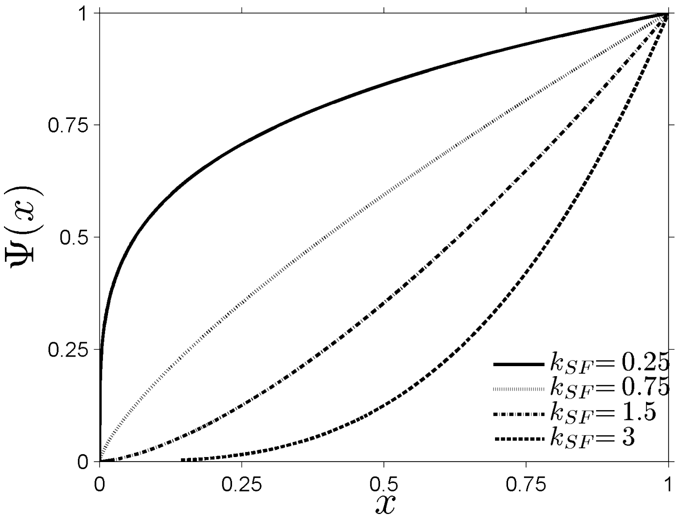



(b)

**Figure 16:** The non linear scaling function $\psi$ and its application to $P_{\mathcal{N}}$ with $\lambda = 0.5$ and $k_{SF} = 0.5$

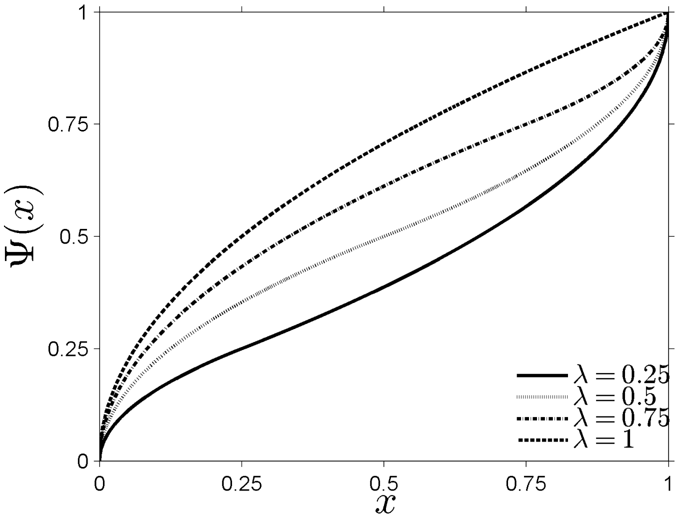



(a) $\lambda = 0.5$

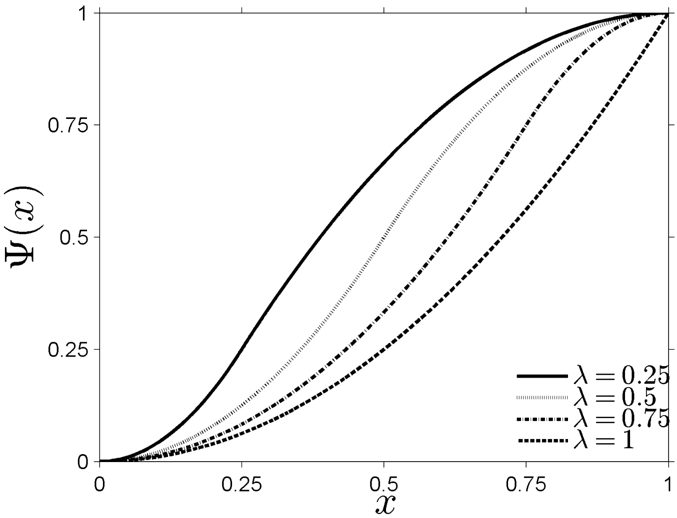



(b) $\lambda = 1$

**Figure 17:** Evaluations of $\psi$ with $k_{SF} = \{0.25, 0.75, 1.5, 3\}$ and (a) $\lambda = 0.5$ and (b) $\lambda = 1$




(a) $k_{SF} = 0.5$




(b) $k_{SF} = 2$

**Figure 18:** Evaluations of $\psi$ with $\lambda = \{0.25, 0.5, 0.75, 1\}$ and (a) $k_{SF} = 0.5$ and (b) $k_{SF} = 2$

We recall that Definitions 22 – 27 introduced in Chapter 2 provide a detailed taxonomy of fuzzy modifiers. To sum up, a fuzzy modifier may be either inclusive or not. Inclusive modifiers may be either expansive or restrictive. Moreover, depending on its formulation, a fuzzy modifier may be decomposed into a pre- and post-modifier. Further subclasses of decomposable fuzzy modifiers are those of pure pre- and post-modifiers.

Various preliminary versions of the modifiers introduced in the following Sections have been defined in our earlier papers (BLM06a; BLM06b; BLMS08; BLM08).

Other remarkable previous work about fuzzy modifiers can be found in (CK00; Coc01; BS01; SWK01; HHN02; CK02; BS03). Recently, an approach similar to the core-width and the support-width modifiers has been proposed in (BDHA+07).

### 3.2.1 Core-Position Modifier

The core-position modifier acts on the core of a fuzzy set, shifting its position within the support while maintaining the original width. The effect of its application produces a shift of the center of mass of fuzzy sets. This shift can, for instance, model the drift of a sensor, which moves the overall distribution of the measures from an initial fixed point to another.

**Definition 37 (Core-position modifier)** *Given a trapezoidal fuzzy set $A(x; sl, cl, cu, su)$ defined on the universe of discourse $\mathcal{U}$, the core-position modifier $m_{CP} : \mathcal{F}(\mathcal{U}) \rightarrow \mathcal{F}(\mathcal{U})$ is given by*

$$A'(x; sl', cl', cu', su') = m_{CP}(A(x; sl, cl, cu, su)), \tag{3.4}$$

*where*

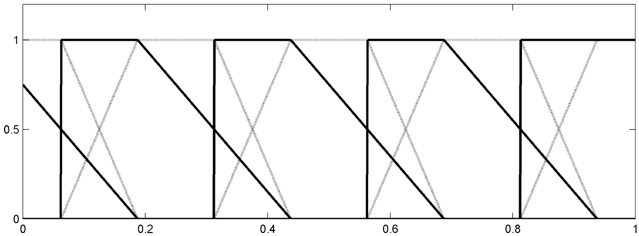
$$sl' = sl, \tag{3.5}$$

$$cl' = \begin{cases} cl - (sl - cl)k_{CP} & \text{if } k_{CP} < 0 \\ cl + (su - cu)k_{CP} & \text{if } k_{CP} \geq 0, \end{cases} \tag{3.6}$$

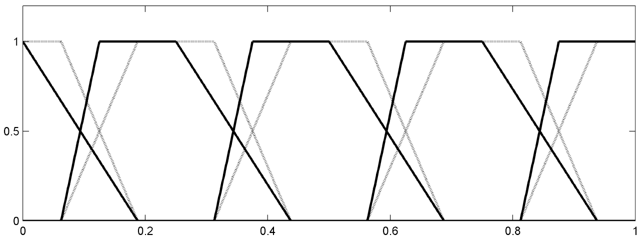
$$cu' = \begin{cases} cu - (sl - cl)k_{CP} & \text{if } k_{CP} < 0 \\ cu + (su - cu)k_{CP} & \text{if } k_{CP} \geq 0, \end{cases} \tag{3.7}$$

$$su' = su, \tag{3.8}$$

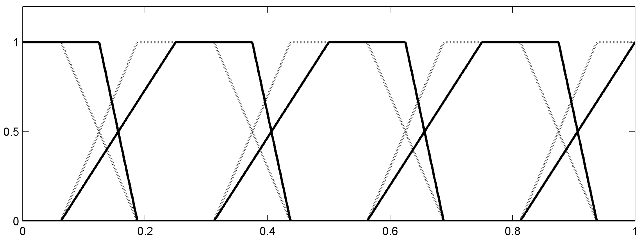
*with $k_{CP} \in [-1, 1]$.*



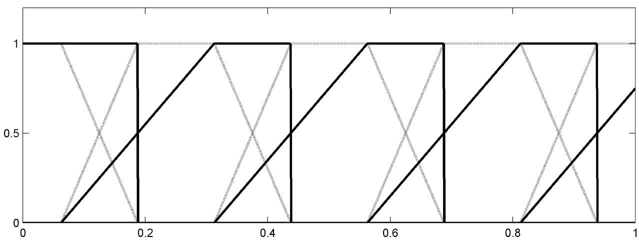

(a) $k_{CP} = -1$




(b) $k_{CP} = -0.5$

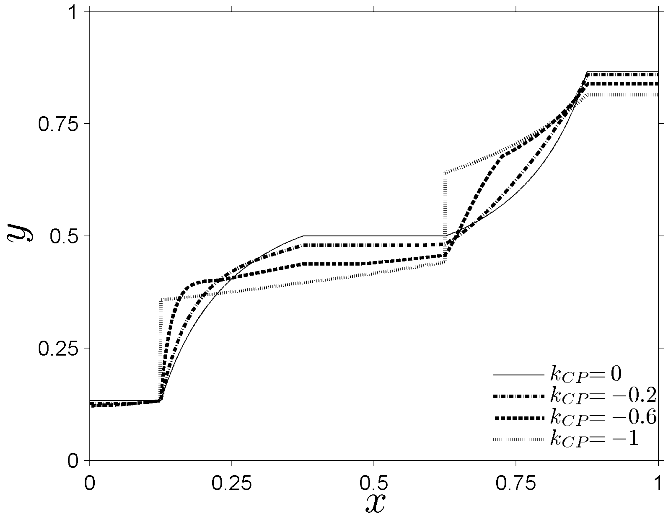



(c) $k_{CP} = 0.5$

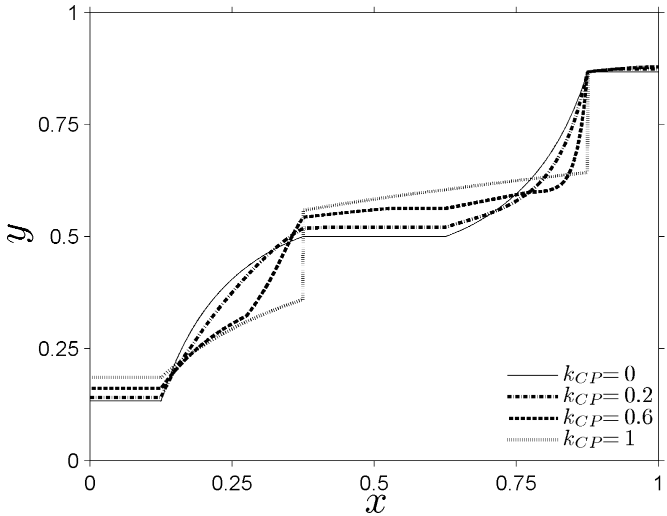



(d) $k_{CP} = 1$

**Figure 19:** Application of the core-position modifier to $P_{\mathcal{N}}$




(a) $k_{CP} \leq 0$




(b) $k_{CP} \geq 0$

**Figure 20:** Effects produced by the core-position modifier when applied to a simple SISO Mamdani-type FRBS

As it can be trivially observed in Equations 3.5 and 3.8, the bounds of the support are not changed by the application of the modifier, while parameter $k_{CP}$ determines the intensity of the left shift ($k_{CP} < 0$) or right shift ($k_{CP} > 0$) of the core. Null values of $k_{CP}$ do not modify the original fuzzy set. The core-position modifier is a pure pre-modifier and it is not inclusive.

Figure 19 shows sample applications of the core-position modifier to the fuzzy partition $P_{\mathcal{N}}$, with $k_{CP} = \{-1, -0.5, 0.5, 1\}$.

To better understand the effects of the application of the core-position modifier to an FRBS, we show a very simple application. Let us consider a single-input single-output (SISO) Mamdani-type FRBS, with three uniformly distributed MFs in $\mathcal{N}$, both for the input and for the output. The RB is composed of three simple rules, all in the form

$$\text{if } x \text{ is } A_i \text{ then } y \text{ is } A_i.$$

We applied the core-position modifier with different values of $k_{CP}$ on both the input and output variables to highlight the effects on the relation identified by the FRBS. Figure 20 shows the results of this sample application. We can easily observe how the shift of the center of mass produced by the application of the modifier reflects into a shift of the breakpoints of the overall input-output relation.

### 3.2.2 Core-Width Modifier

The core-width modifier acts on the core of a fuzzy set, dilating or shrinking its width within the support. The effect of the modifier is to increase or decrease the number of points that belong to the fuzzy set with full degree. Consequently, the modifier affects the uncertainty modeled by the fuzzy set. A side effect of the application of this modifier on all the fuzzy sets of a fuzzy partition is the change of the level of coverage of the partition.

**Definition 38 (Core-width modifier)** *Given a trapezoidal fuzzy set* $A(x; sl, cl, cu, su)$ *defined on the universe of discourse* $\mathcal{U}$*, the core-width modifier* $m_{CW} : \mathcal{F}(\mathcal{U}) \to \mathcal{F}(\mathcal{U})$ *is given by*

$$A'(x; sl', cl', cu', su') = m_{CW}(A(x; sl, cl, cu, su)), \tag{3.9}$$

*where*

$$\begin{aligned}
sl' &= sl, && (3.10)\\
cl' &= \begin{cases} cl + w(sl - cl)k_{CW} & \text{if } k_{CW} < 0 \\ cl + (sl - cl)k_{CW} & \text{if } k_{CW} \geq 0, \end{cases} && (3.11)\\
cu' &= \begin{cases} cu + w(su - cu)k_{CW} & \text{if } k_{CW} < 0 \\ cu + (su - cu)k_{CW} & \text{if } k_{CW} \geq 0, \end{cases} && (3.12)\\
su' &= su, && (3.13)
\end{aligned}$$

*with* $k_{CW} \in [-1, 1]$ *and*

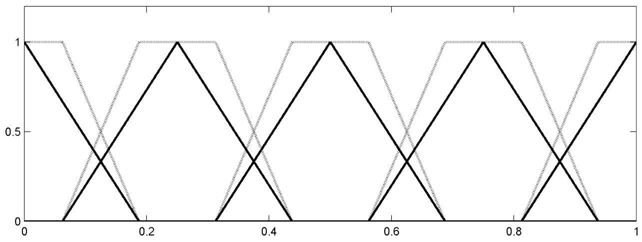
$$w = \frac{cu - cl}{cl - sl + su - cu}. \tag{3.14}$$

As in the case of the core-position modifier, the bounds of the support are not modified, and parameter $k_{CW} \in [-1, 1]$ determines the intensity of dilation ($k_{CW} > 0$) or shrinking ($k_{CW} < 0$) of the core. Null values of $k_{CW}$ do not modify the original fuzzy set. The core-width modifier is an inclusive pure pre-modifier, restrictive when $k_{CW}$ is negative and expansive when $k_{CW}$ is positive.

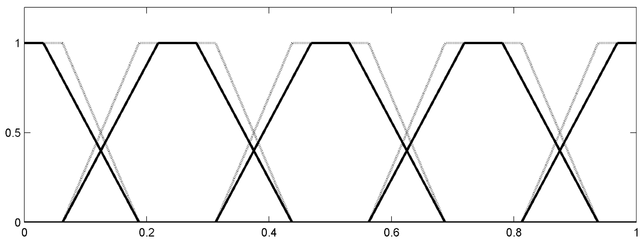
We remark that, when the modifier is applied to trapezoidal MFs with $k_{CW} = -1$, the fuzzy sets degenerate into triangular MFs. On the other hand, when $k_{CW} = 1$, the fuzzy sets degenerate into crisp sets, thus arising some interpretability issues because the MFs overlap too much.

Figure 21 shows sample applications of the core-width modifier on the fuzzy partition $P_{\mathcal{N}}$, with $k_{CW} = \{-1, -0.5, 0.25, 0.5\}$. As stated above, the application of the core-width modifier might produce some interpretability problems. Indeed, when applied with $k_{CW} > 0.5$ on $P_{\mathcal{N}}$, the level of coverage of the modified partition gets to 1.

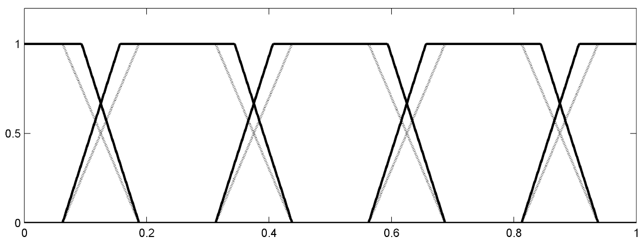
Figure 22 shows the effects of the application of the core-width modifier to the simple SISO described above with different positive values of $k_{CW}$ (in this trivial example, negative values of $k_{CW}$ do not produce meaningful effects on the input-output relation of the FRBS). As stated above, the application of the modifier increases or decreases the uncertainty (i.e., the stepwise behavior) of each single fuzzy set in the partition. Indeed, smaller values of $k_{CW}$ produce a smooth input-output relation, whilst when $k_{CW} = 1$ the curve becomes completely crispified.



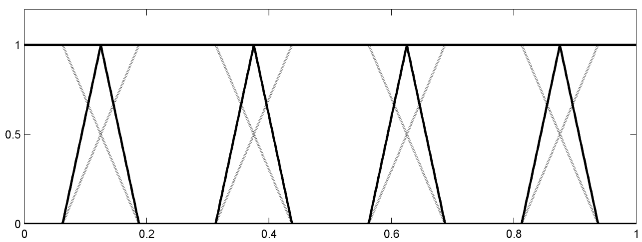
(a) $k_{CW} = -1$



(b) $k_{CW} = -0.5$



(c) $k_{CW} = 0.25$



(d) $k_{CW} = 0.5$

**Figure 21:** Application of the core-width modifier to $P_{\mathcal{N}}$

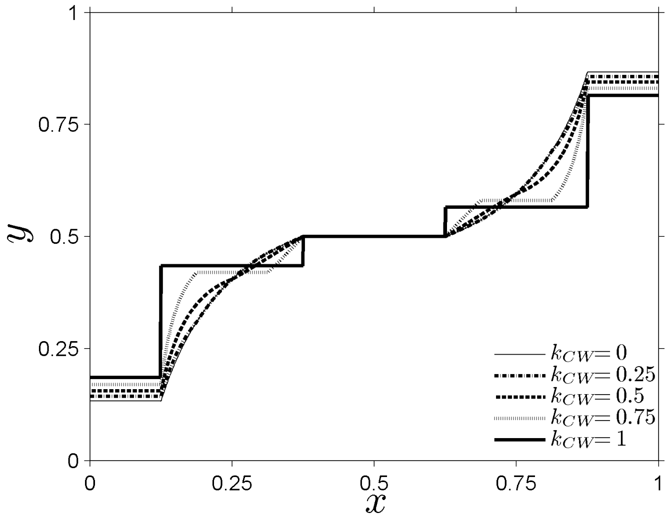



**Figure 22:** Effects produced by the core-width modifier when applied to a simple SISO Mamdani-type FRBS

### 3.2.3 Support-Width Modifier

The support-width modifier acts on both the support and the core of a fuzzy set, scaling their widths with respect to the center of the support and preserving the ratio between the widths of the core and the support. The application of the modifier, therefore, increases or reduces the number of points belonging to the fuzzy set to some extent.

The support-width modifier is inspired by the concentration hedge *CON* introduced in (BS03). Like the core-width modifier, the support-width modifier might change the level of coverage of a partition to a high extent.

**Definition 39 (Support-width modifier)** *Given a trapezoidal fuzzy set* $A(x; sl, cl, cu, su)$ *defined on the universe of discourse* $\mathcal{U}$*, the support-width modifier* $m_{SW} : \mathcal{F}(\mathcal{U}) \to \mathcal{F}(\mathcal{U})$ *is given by*

$$A'(x; sl', cl', cu', su') = m_{SW}(A(x; sl, cl, cu, su)), \tag{3.15}$$

*where*

$$
\begin{aligned}
sl' &= sm + (sl - sm)k_{SW}, & (3.16)\\
cl' &= sm + (cl - sm)k_{SW}, & (3.17)\\
cu' &= sm + (cu - sm)k_{SW}, & (3.18)\\
su' &= sm + (su - sm)k_{SW}, & (3.19)
\end{aligned}
$$

*with* $k_{SW} > 0$ *and*

$$
sm = \frac{sl + su}{2}. \quad (3.20)
$$

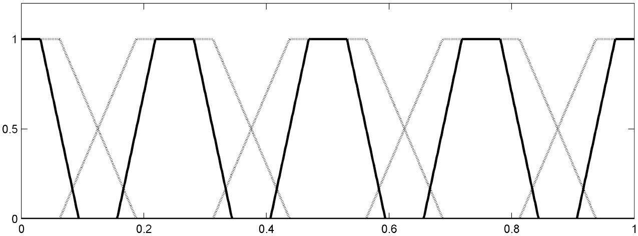
Parameter $k_{SW}$ determines the negative ($k_{SW} < 1$) or positive ($k_{SW} > 1$) scaling of the fuzzy set. When $k_{SW} = 1$, no change is determined. The support-width modifier is an inclusive pure pre-modifier, restrictive when $k_{SW} < 1$ and expansive when $k_{CW} > 1$.

Figure 23 shows a sample application of the support-width modifier on the fuzzy partition $P_{\mathcal{N}}$, with $k_{SW} = \{0.5, 0.75, 1.5, 2\}$. Since the application of the operator can reduce the support of all the MFs of a partition, coverage problems might occur, i.e., some parts of the universe might be covered by no fuzzy set, as in Figure 23(a). This is a well-known problem in fuzzy modeling, especially for the input linguistic variables (dO99), which can be simply resolved by restricting the range of possible values for parameter $k_{SW}$.

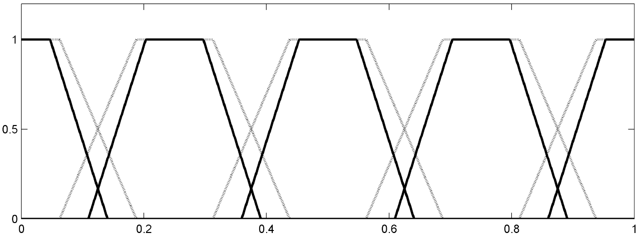
Figure 24 shows the effects of the application of the support-width modifier to the simple SISO with different values of $k_{SW}$. As expected, the larger the value of $k_{SW}$, the more the FRBS gives an uncertain (i.e., averaged) output. On the other hand, the smaller the values of $k_{SW}$, the more crispified is the output, i.e., the more the input-output relation has a stepwise behavior. Note that, differently from the core-width modifier, which affects uncertainty of each fuzzy set, the support-width modifier modifies the smoothness of the overall curve.

### 3.2.4 Generalized Positively Modifier

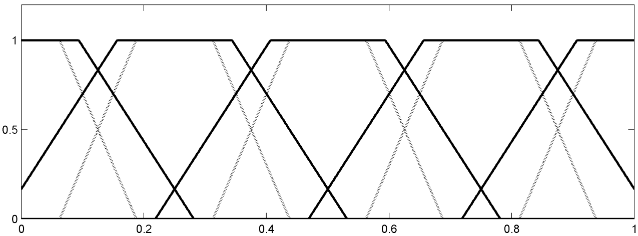
The three modifiers introduced so far act on the core and the support of the MFs, without modifying the type of function that defines the membership degrees of the boundary elements. In the fuzzy partition $P_{\mathcal{N}}$,



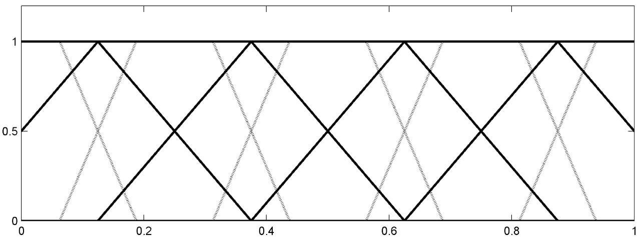

(a) $k_{SW} = 0.5$




(b) $k_{SW} = 0.75$




(c) $k_{SW} = 1.5$




(d) $k_{SW} = 2$

**Figure 23:** Application of the support-width modifier to $P_{\mathcal{N}}$

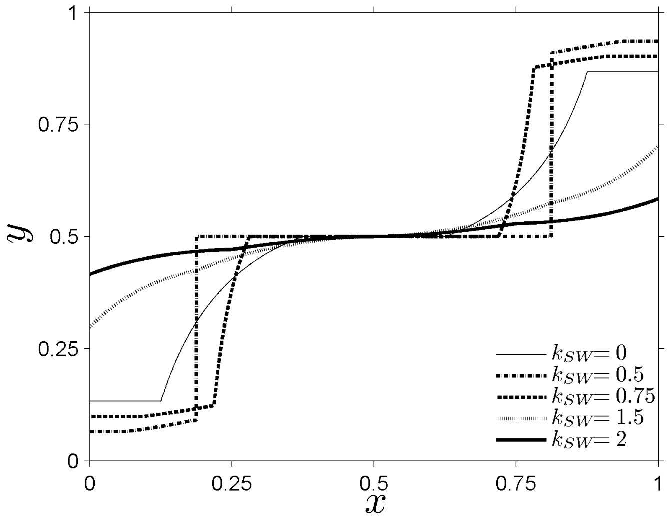



**Figure 24:** Effects produced by the support-width modifier when applied to a simple SISO Mamdani-type FRBS

this function is linear. As shown in Figures 19, 21, and 23, the function remains linear also after applying the three modifiers.

The linear type of function, however, might not be the most appropriate to correctly model the effects of the context. Thus, we need to introduce a fuzzy modifier that allows adapting the shape of the MFs. To this aim, we generalize the linguistic hedge *positively* defined in the literature as (Zad73)

$$m_P(A(x)) : \mathcal{F}(\mathcal{U}) \rightarrow \mathcal{F}(\mathcal{U}) = \begin{cases} 2A^2(x) & \text{if } A(x) < 0.5 \\ 1 - 2[1 - A(x)]^2 & \text{if } A(x) \geq \theta. \end{cases} \tag{3.21}$$

In (SWK01), the modifier has been modified by introducing a parameter $k_{GP}$ for customizing the contrast intensification. We further generalize the modifier by inserting a parameter $\theta \in [0, 1]$ to control the membership value in which the MF changes concavity.

**Definition 40 (Generalized positively modifier)** *Given a fuzzy set $A$ defined on the universe of discourse $\mathcal{U}$, the generalized modifier $m_{GP} : \mathcal{F}(\mathcal{U}) \rightarrow \mathcal{F}(\mathcal{U})$ is defined as*

$$A'(x) = m_{GP}\left(A\left(x\right)\right), \tag{3.22}$$

*where*

$$A'(x) = \begin{cases} \theta^{1-k_{GP}} A^{k_{GP}}(x) & \text{if } A(x) < \theta \\ 1 - (1-\theta)^{1-k_{GP}}[1 - A(x)]^{k_{GP}} & \text{if } A(x) \geq \theta, \end{cases} \tag{3.23}$$

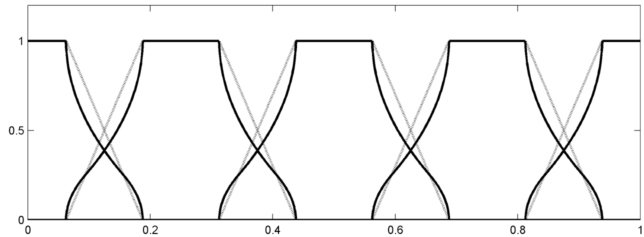
*with* $k_{GP} > 0$ *and* $\theta \in [0, 1]$.

This modifier is extremely powerful: it can generate a large number of different shapes, using a compact notation and just two parameters. Further, with particular values of $k_{GP}$ and $\theta$, it can reproduce the effects of the original linguistic hedges *positively* ($\theta = 0.5$, $k_{GP} = 2$) and of other three linguistic hedges introduced by Zadeh (*very*, *more-or-less*, and *negatively*).

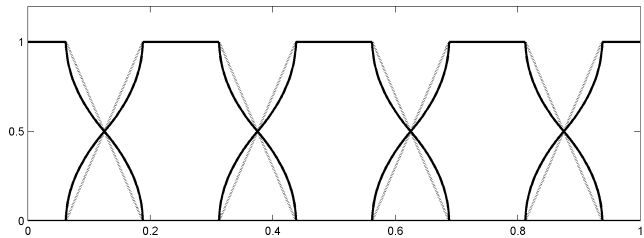
We remark that, differently from the other modifiers defined in this Chapter, the generalized positively modifier is a pure post-modifier, since it applies to the membership degrees of the fuzzy set rather than to the universe of discourse. The generalized positively modifier is not inclusive, except for the special cases $\theta = 0$ and $\theta = 1$: in these cases, it is either restrictive (if $k_{GP} < 1$) or expansive (if $k_{GP} > 1$). When $k_{GP} \to 0$ or $k_{GP} \gg 1$, the modifier can generate strange shapes which, eventually, may degenerate into crisp or singleton MFs.

Figures 25 – 26 show sample applications of the generalized positively modifier on the fuzzy partition $P_{\mathcal{N}}$, with $k_{GP} = \{0.5, 2.5\}$ and $\theta = \{0.25, 0.5, 1\}$.

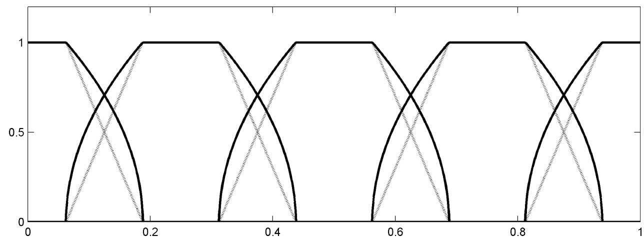
Figure 27 shows the effects of the application of the generalized positively modifier to the simple SISO with different values of $k_{GP}$, namely $k_{GP} = \{0.25, 0.5, 1, 2, 4\}$, and with $\theta = 0.25$ and $\theta = 0.75$. The application of this modifier enhances the smoothness (thanks to the contrast intensification) and, concurrently, can move the breakpoints of the input-output relation (thanks to the tuning of $\theta$). As expected, the larger the value of $k_{SW}$, the more the FRBS gives an uncertain (i.e., averaged) output. On the other hand, the smaller the values of $k_{SW}$, the more crispified is the output, i.e., the more the input-output relation has a stepwise behavior.




(a) $k_{GP} = 0.5, \theta = 0.25$

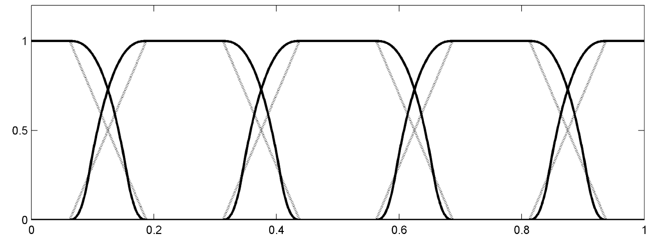



(b) $k_{GP} = 0.5, \theta = 0.5$

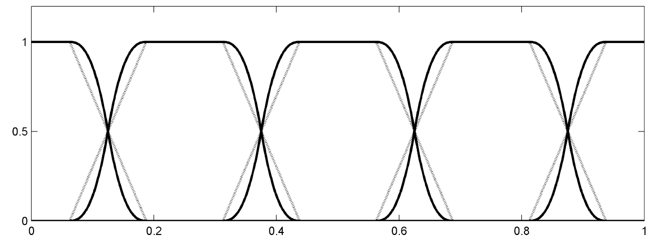



(c) $k_{GP} = 0.5, \theta = 1$

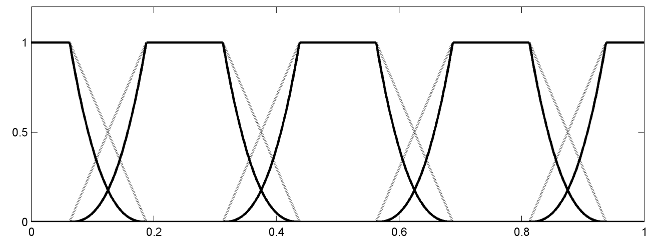
**Figure 25:** Application of the generalized positively to $P_{\mathcal{N}}$




(a) $k_{GP} = 2.5, \theta = 0.25$

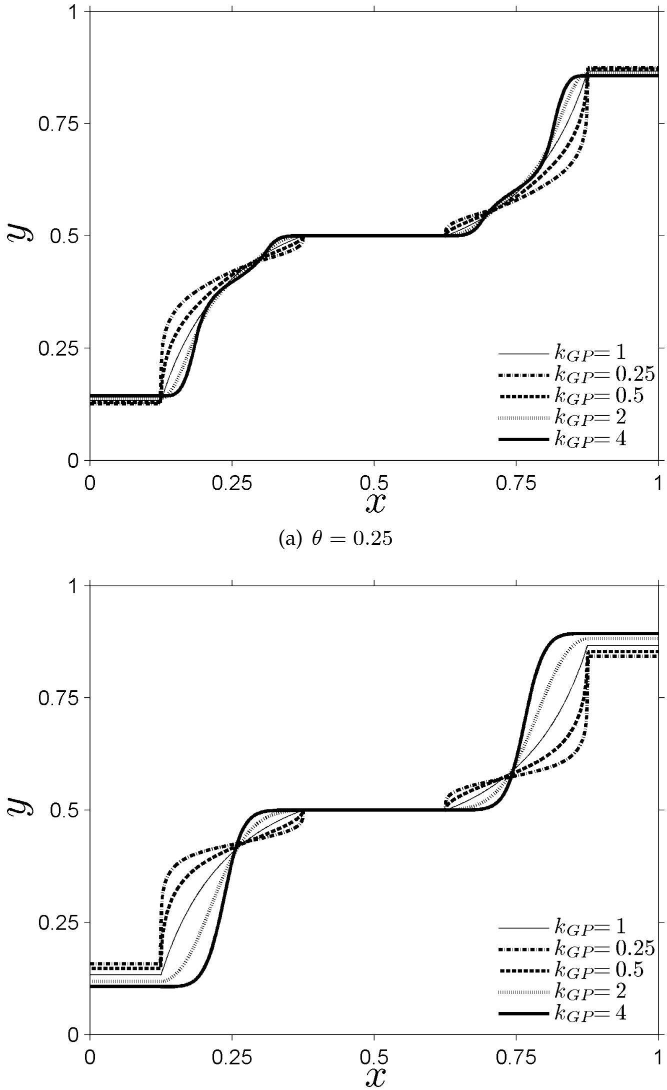



(b) $k_{GP} = 2.5, \theta = 0.5$




(c) $k_{GP} = 2.5, \theta = 1$

**Figure 26:** Application of the generalized positively to $P_{\mathcal{N}}$




(a) $\theta = 0.25$



(b) $\theta = 0.75$

**Figure 27:** Effects produced by the generalized positively modifier when applied to a simple SISO Mamdani-type FRBS

### 3.2.5 Orthogonality of Fuzzy Modifiers

An important mathematical property of the fuzzy modifiers introduced in the previous Sections is that they are orthogonal with respect to each other, i.e., their composition is commutative and associative and, hence, the order in which they are applied to a fuzzy partition is not relevant.

To prove the orthogonality of the fuzzy modifiers, we have to check whether all the possible combinations in different orderings produce the same modified fuzzy set. Given the complex and/or conditional formulation of most of the modifiers, the computation required for this proof is extremely hard. However, such difficult task can be simplified by the following observations.

- The generalized positively modifier is a pure post-modifier which changes the membership degree of boundary elements without affecting the breakpoints. On the contrary, the other modifiers are pure pre-modifiers which act on the breakpoints of the trapezoidal MFs. Hence, we do not need to include the generalized positively modifier in the test, because its application does not interfere with the application of the other modifiers. Formally speaking, it can be easily verified that

$$(A' = m_{GP}(A)) \Rightarrow \begin{cases} sl' & = & sl \\ cl' & = & cl \\ cu' & = & cu \\ su' & = & su. \end{cases} \tag{3.24}$$

- Not all of the combinations of compositions have to be checked for equivalence. Indeed, by carefully choosing a minimal set of tests and by exploiting the commutativity and associativity properties, we can avoid to assess some equivalences.

Given a trapezoidal fuzzy set $A(x; sl, cl, cu, su)$, the chosen minimal set of equations to check is composed of

$$
\begin{aligned}
m_{CP}\left(m_{CW}\left(A\right)\right) &= m_{CW}\left(m_{CP}\left(A\right)\right), && (3.25)\\
m_{CP}\left(m_{SW}\left(A\right)\right) &= m_{SW}\left(m_{CP}\left(A\right)\right), && (3.26)\\
m_{CW}\left(m_{SW}\left(A\right)\right) &= m_{SW}\left(m_{CW}\left(A\right)\right), && (3.27)\\
m_{SW}\left(m_{CP}\left(m_{CW}\left(A\right)\right)\right) &= m_{CP}\left(m_{CW}\left(m_{SW}\left(A\right)\right)\right), && (3.28)\\
m_{CW}\left(m_{CP}\left(m_{SW}\left(A\right)\right)\right) &= m_{CP}\left(m_{SW}\left(m_{CW}\left(A\right)\right)\right), && (3.29)\\
m_{CP}\left(m_{CW}\left(m_{SW}\left(A\right)\right)\right) &= m_{SW}\left(m_{CP}\left(m_{CW}\left(A\right)\right)\right). && (3.30)
\end{aligned}
$$

Although the complete set of tests should comprise twelve more expressions, the minimal set of Equations 3.25 – 3.30 can be used to check all possible combinations. For instance, if Equations 3.25 and 3.28 are verified, it follows that

$$
m_{SW}\left(m_{CW}\left(m_{CP}\left(A\right)\right)\right) = m_{CP}\left(m_{CW}\left(m_{SW}\left(A\right)\right)\right), \qquad (3.31)
$$

is verified as well.

Equations 3.25 – 3.30 were solved with the aid of the *Matlab Symbolic Math Toolbox* (Mat06), and all the equivalences were correctly verified. The complete report of the analytical steps of the computation is out of the scope of the current work. For the sake of completeness, in the following we show an example of some steps, performed to assess Equation 3.25 in the case $k_{CP} > 0$ and $k_{CW} > 0$.

$$
\begin{aligned}
A'(x; sl', cl', cu', su') &= m_{CP}\left(m_{CW}\left(sl\left(x; sl, cl, cu, su\right)\right)\right)\\
A''(x; sl'', cl'', cu'', su'') &= m_{CW}\left(m_{CP}\left(sl\left(x; sl, cl, cu, su\right)\right)\right)
\end{aligned}
$$

$$
\begin{cases}
sl' = sl\\
cl' = (su - cu - (su - cu)k_{CW})k_{CP} + cl - (cl - sl)k_{CW}\\
cu' = cu + (su - cu)\,k_{CW} + (su - cu - (su - cu)\,k_{CW})\,k_{CP}\\
su' = su,
\end{cases}
$$

$$
\begin{cases}
sl'' = sl\\
cl'' = cl + (su - cu)\,k_{CP} - (cl + (su - cu)\,k_{CP} - sl)\,k_{CW}\\
cu'' = cu + (su - cu)\,k_{CP} + (su - cu - (su - cu)\,k_{CP})\,k_{CW}\\
su'' = su.
\end{cases}
$$

It can be trivially verified that $cl' = cl''$ and $cu' = cu''$, and, therefore, Equation 3.25 is verified for $k_{CP} > 0$ and $k_{CW} > 0$. Similar tests have been repeated for $k_{CP} > 0$ and $k_{CW} < 0$, $k_{CP} < 0$ and $k_{CW} > 0$, and $k_{CP} < 0$ and $k_{CW} < 0$ to check the validity of the equivalence in Equation 3.25.

# Chapter 4

# Interpretability Issues in Context Adaptation of Fuzzy Systems

As we have stated in the previous Chapters, CA of an FRBS consists in tuning the fuzzy sets corresponding to the linguistic terms used in the RB. The approach proposed in this thesis exploits a non linear scaling function and four fuzzy modifiers that are able to reproduce the effects of the context on each considered fuzzy partition.

Even though the operators are designed to reproduce the designer's choice when modeling linguistic concepts, their combined effects may sometimes lead to the generation of a fuzzy partition which is poorly human-readable. Usually, this happens when we exploit the augmented flexibility provided by fuzzy modifiers to achieve good performances of context-adapted systems. As it is well-known in the field of FRBS learning from data, interpretability and accuracy are conflicting objectives and, therefore, identification algorithms must carefully take their trade-off into account (CdJH+03; CGH+04; IN07; Her08).

In this Chapter, we first define the relation between interpretability of FRBSs and CA. Then, we develop two approaches to preserve integrity of fuzzy partitions based on the evaluation of crossing-points and fuzzy ordering relations, respectively.

## 4.1 Interpretability of FRBSs

The issue of balancing interpretability and accuracy has been widely addressed in the field of FRBS generation from data (dO99; CCHM03). Interpretability of an FRBS cannot be defined in a unique way and, therefore, several different approaches have been developed in the literature. Related work has been surveyed in (Gui01; CCHM03; CdJH$^+$03; Men04).

Typically, we distinguish between *interpretability of fuzzy partitions* (i.e., of the DB), sometimes also called integrity or transparency, and *interpretability of rules* (i.e., of the RB), also known as complexity.

Complexity is often defined in terms of simple measures, such as number of rules in the RB and number of linguistic terms in their antecedents.

On the other hand, integrity depends on some intrinsic properties of the fuzzy partitions contained in the DB, such as, for instance, coverage, ordering, continuity, and normality, which may be difficult to measure directly. Indeed, in the majority of existing approaches, interpretability of fuzzy partitions is typically measured indirectly by exploiting simple similarity indices that cannot always completely capture the actual semantics of the partitions (SBKvNL98; SR00; MCFB04; ZYYLY$^+$06).

For instance, a common way to measure interpretability is to compute the fuzzy extension of the Jaccard similarity index (CS02) for each couple of fuzzy sets in the partition.

**Definition 41** *Given two fuzzy sets $A_1, A_2 \in \mathcal{F}(\mathcal{U})$, the Jaccard index is defined as*

$$S_J(A_1, A_2) = \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|}. \tag{4.1}$$

One of the most agreed definition of interpretability of fuzzy partition can be found in (dO99). Here, de Oliveira stated that a fuzzy partition is interpretable if it satisfies all of the following properties.

1. The partition should comprise a "moderate" number of fuzzy sets.

2. The fuzzy sets in the partition should all be normal, i.e., for each fuzzy set there must exist at least one point with membership degree equal to 1 (see also Definition 16).

3. Each couple of fuzzy sets should be distinguishable enough, so that there is no couple of fuzzy sets that represents pretty much the same concept.

4. The overall universe of discourse should be strictly covered, i.e., each point of the universe should belong to at least one fuzzy set with a membership degree greater than a given reasonable threshold.

Further, to make a partition interpretable, de Oliveira suggested the use of a linguistic term to represent the fuzzy number *zero*, which can be used as a reference to understand the meaning of the other fuzzy sets. The natural *zero* is commonly adopted in control applications, but it is not suited for many modeling and regression examples which may not involve a symmetric quantization of the universe of discourse.

Some researchers, like Mencar et al. (MCFB04), suggest to take other mathematical properties of fuzzy sets into account, e.g., continuity. In this thesis, however, we focus on general features that are independent of the actual universe of discourse on which the linguistic variable is defined.

To give an example of the contrast between interpretable and non interpretable fuzzy partitions, let us consider Figure 28, which shows two partitions composed of five trapezoidal fuzzy sets each. Figure 28(a) reports a sample of a poorly human-readable partition, like the ones that are typically generated by an accuracy-oriented algorithm, whilst Figure 28(b) shows a clearly readable partition with good distinguishability and coverage.

With respect to CA of FRBSs, we recall that, as stated in Section 2.2.2, we start from the assumption that the RB has an universal validity and is properly selected during the knowledge elicitation and/or the abstraction processes. Thus, we assume that the RB is already human-readable and concentrate just on interpretability of the DB.

Besides de Oliveira's requirements, CA of FRBSs forces other constraints on the interpretability of the DB. Indeed, when designing an interpretability index for CA, we should take into account both the three guidelines defined in Section 2.2.2 and the four properties cited above.

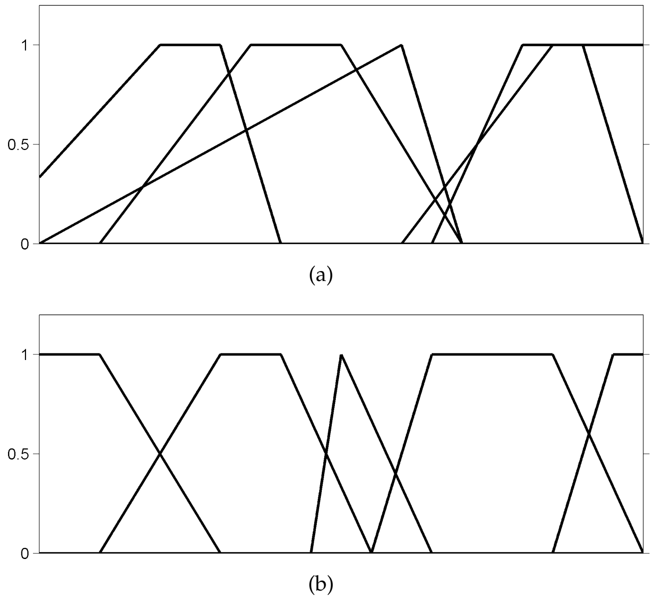



**Figure 28:** Examples of (a) a non interpretable and (b) an interpretable fuzzy partitions composed of five fuzzy sets

The number of linguistic terms of a context-free universal model should be usually low, because the RB is extracted from existing knowledge by elicitation, which is a human-driven process, or abstraction, which can be designed for a proper optimization of the size of the RB (WM92). Hence, if, as stated by guideline 1, the RB is not modified during instantiation, then property 1 (low number of fuzzy sets) is enforced as well. Further, if the operators used for CA do not alter the normality of the context-adapted fuzzy sets, as the novel ones introduced in Chapter 3, property 2 (normality) is verified as well. On the other hand, properties 3 and 4 are not so easily satisfiable.

Indeed, the latter two requirements pose interesting challenges to the designer of a learning method for FRBSs. First, defining a proper metric to measure them with low computational effort is difficult. Second, it is hard to find a crisp threshold for the metric so as to separate good from bad partitions.

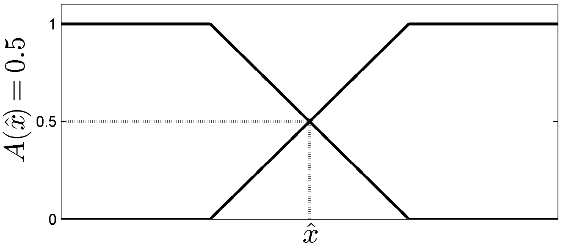
It can be shown that Jaccard's index does not seem to provide means for salient evaluation of ordering of fuzzy partitions. On the other hand, as we detail in Section 4.3, an ordering indices can be used also to successfully evaluate distinguishability and coverage from an interpretability point-of-view.

## 4.2 Preserving Interpretability through Evaluation of Crossing Points

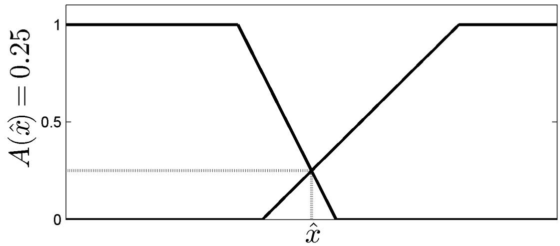
In this Section, we introduce a simple technique to handle interpretability of a fuzzy partition during CA of FRBSs. An approach to address this issue is to consider distinguishability and coverage as conflicting properties, and to use a metric to assess the trade-off between them. As stated above, in previous work this is typically performed by the use of a similarity index, e.g., Jaccard's one. However, similarity indices are affected by two major drawbacks.

First, such an index typically involves the calculation of the cardinality of one or more fuzzy sets. From a computational point-of-view, this is a very expensive task, because it includes integrating over the universe of discourse of the base variable. Second, although similarity can be a good measure of distinguishability, it is not always well suited to assess the actual coverage of a partition. Therefore, by assessing interpretability via a similarity index, we run the risk of spending significant computational resources and obtaining just a rough estimate of the actual properties of the partition.

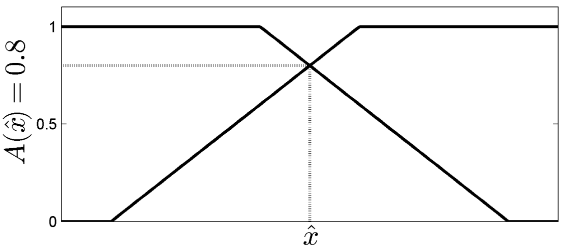
Here, we propose to assess coverage and distinguishability by exploiting the membership values of crossing points between adjacent fuzzy sets. To this aim, let us consider a partition $P = \{A_1, ..., A_i, ..., A_N\}$ which comprises $N$ trapezoidal fuzzy sets. Two *adjacent fuzzy sets* in $P$ are two fuzzy sets $A_i$ and $A_{i+1}$. The *crossing point* $\hat{x}_i$ between $A_i$ and $A_{i+1}$ is defined as the point where the right spread of $A_i$ equals the left spread of $A_{i+1}$. It can be easily observed that the membership value $A_i(\hat{x}_i)$ is a direct measure of the coverage of the partition. We remark that, if either $\mathcal{C}(A_i) \cap \mathcal{C}(A_{i+1}) \neq \emptyset$ or $\mathcal{S}(A_i) \cap \mathcal{S}(A_{i+1}) = \emptyset$, then $\hat{x}_i$ cannot be computed. In such cases, we set $\hat{x}_i = 1$ or $\hat{x}_i = 0$, respectively.




(a)




(b)




(c)

**Figure 29:** Examples of evaluation of crossing points for the assessment of distinguishability and coverage

The appropriate definition of the fuzzy modifiers and the choice of trapezoidal MFs allow us to compute the crossing point of adjacent fuzzy sets exactly and with minimal computational efforts. Indeed, $\hat{x}_i$ is simply the point that solves the equation

$$\begin{aligned} & m_{CP}\left(m_{CW}\left(m_{SW}\left(m_{GP}\left(A\left(x; sl_i, cl_i, cu_i, su_i\right)\right)\right)\right)\right) = \\ & m_{CP}\left(m_{CW}\left(m_{SW}\left(m_{GP}\left(A\left(x; sl_{i+1}, cl_{i+1}, cu_{i+1}, su_{i+1}\right)\right)\right)\right)\right). \end{aligned} \tag{4.2}$$

We remark that the order of application of the modifiers is irrelevant, as proven in Section 3.2.5. Also, Equation 4.2 can be simplified analytically and expressed in terms of the parameters of the MFs and of the modifiers, as

$$\hat{x}_i = f_{XP}(cu_i, su_i, sl_{i+1}, cl_{i+1}, k_{CP}, k_{CW}, k_{SW}, k_{GP}, \theta). \tag{4.3}$$

We skip the long formulation of $f_{XP}$ because it is not of interest for the current analysis.

As stated above, $A_i(\hat{x}_i)$ is a direct measure of the coverage level achieved by the couple of fuzzy sets $(A_i, A_{i+1})$. Hence, assessing the crossing points of all the couples of adjacent fuzzy sets, we can compute the overall level of coverage of a partition. On the other hand, $A_i(\hat{x}_i)$ is also a good estimator of the distinguishability of $A_i$ and $A_{i+1}$. Indeed, as shown in the examples reported in Figure 29, high values of $A_i(\hat{x}_i)$ typically also correspond to poor distinguishability.

Taking all these observations into account, we can define a *penalty index* for fuzzy partitions that measures how many couples of adjacent fuzzy sets violates the distinguishability and the coverage constraints expressed in terms of a range of allowed values for $A_i(\hat{x}_i)$.

**Definition 42 (Interpretability index $\Phi_{XP}$)** *Given a partition $P = \{A_1, ..., A_i, ..., A_N\}$ consisting of N fuzzy sets, the interpretability index $\Phi_{XP}$ is defined as*

$$\Phi_{XP}(P) = \sum_{i=1}^{N-1} \phi_i, \tag{4.4}$$

*with*

$$\phi_i = \begin{cases} 1 & \text{if } A_i(\hat{x}_i) < \epsilon_{\min} \vee A_i(\hat{x}_i) > \epsilon_{\max} \\ 0 & \text{otherwise,} \end{cases} \tag{4.5}$$

*where* $\epsilon_{\min}, \epsilon_{\max} \in [0,1]$, $\epsilon_{\min} < \epsilon_{\max}$, *are the thresholds for coverage and distinguishability penalties, respectively.*

Typical values for $\epsilon_{\min}$ and $\epsilon_{\max}$ are 0.25 and 0.75, respectively.

## 4.3 Preserving Interpretability through Fuzzy Ordering Relations

In the previous Section, we have introduced a simple technique to assess coverage and distinguishability of fuzzy partition during the instantiation process performed in CA of FRBSs. However, since in CA we also require to preserve ordering of the original fuzzy partition, we choose to evaluate the interpretability by extending an ordering index. Indeed, it is extremely difficult to derive useful information about ordering from traditional interpretability indices. Therefore, in this Section, we introduce a novel index that, exploiting a *fuzzy ordering relation*, explicitly takes coverage and distinguishability into account, in attempt to reproduce the interpretability perceived by humans and described by de Oliveira in (dO99).

The definition of our index reflects the following observation: humans associate a semantic ordering with the linguistic terms used as values of a linguistic variable. This ordering has, by and large, universal acceptance and has to be observed by the fuzzy sets used to define the meaning of the linguistic terms employed by the system.

A further condition for interpretability is that fuzzy sets should be made distinguishable from each other so as to preserve distinction between linguistic terms. Indeed, humans associate completely different meanings with different linguistic terms, and these differences are more marked for linguistic terms which are semantically far. For instance, the distinction between *low* and *medium* is less marked than between *low* and *high*.

Finally, the universe should be covered, that is, there should not exist members of the universe which are represented by no linguistic term. The index we propose in this paper considers explicitly these three aspects of interpretability.

### 4.3.1 Fuzzy Sets Ordering

The ordering of fuzzy sets has been widely addressed in the literature (Wan97; WK01a; CS02). For the sake of simplicity, in the following we will restrict our analysis to the case of fuzzy numbers (see Definition 20), which are commonly employed in Mamdani-type FRBSs.

Generally, given two fuzzy numbers $A_1$ and $A_2$ defined on the universe of discourse $\mathbb{R}$, it may be difficult to determine whether and how $A_1$ and $A_2$ are ordered.

A trivial approach is to consider basic features of each fuzzy set, such as the position of modal values or the lower (upper) bound of the support, and to define the ordering of fuzzy sets based on the ordering of these metrics. This approach performs correctly when fuzzy numbers are clearly and intuitively ordered, but it may lead to counter-intuitive results in the general case. Such disputed cases can be handled, for instance, by defining a partial order rather than a total one. A partial order of fuzzy numbers is simply obtained by extending the interval-set theory to $\alpha$-cuts of fuzzy sets (Ros04).

Let us consider the eight case studies of couples of fuzzy sets shown in Figure 30 (case 4 is taken from an example in (WK01a)). Clearly, the ordering of cases 1-3 is easily identifiable, while the other cases are disputable. As stated above, we could either define a $\mathcal{C}$-partial order of $A_1$ and $A_2$ based on the partial order of their cores $\mathcal{C}(A_1)$ and $\mathcal{C}(A_2)$ (i.e., their 1-cuts) or, similarly, an $\mathcal{S}$-partial order based on the partial order of their supports $\mathcal{S}(A_1)$ and $\mathcal{S}(A_2)$ (i.e., their $0^+$-cuts). For instance, in case 8, the $\mathcal{C}$-partial order returns $A_1 \leq A_2$, whilst the $\mathcal{S}$-partial order gives $A_2 \leq A_1$. In case 7, the $\mathcal{C}$-partial order returns $A_1 \leq A_2$, whilst the $\mathcal{S}$-partial order cannot be computed. In cases 4 and 5, neither the $\mathcal{C}$- nor the $\mathcal{S}$-partial orders can be computed. Hence, $\alpha$-cut-based partial orders do not seem a proper solution to determine the ordering of fuzzy sets, because, besides not always being able to compute the actual order, they depend too much on the choice of the $\alpha$-cut.

In (WK01a; WK01b), Wang and Kerre reviewed a number of ordering approaches, and classified them into three categories.

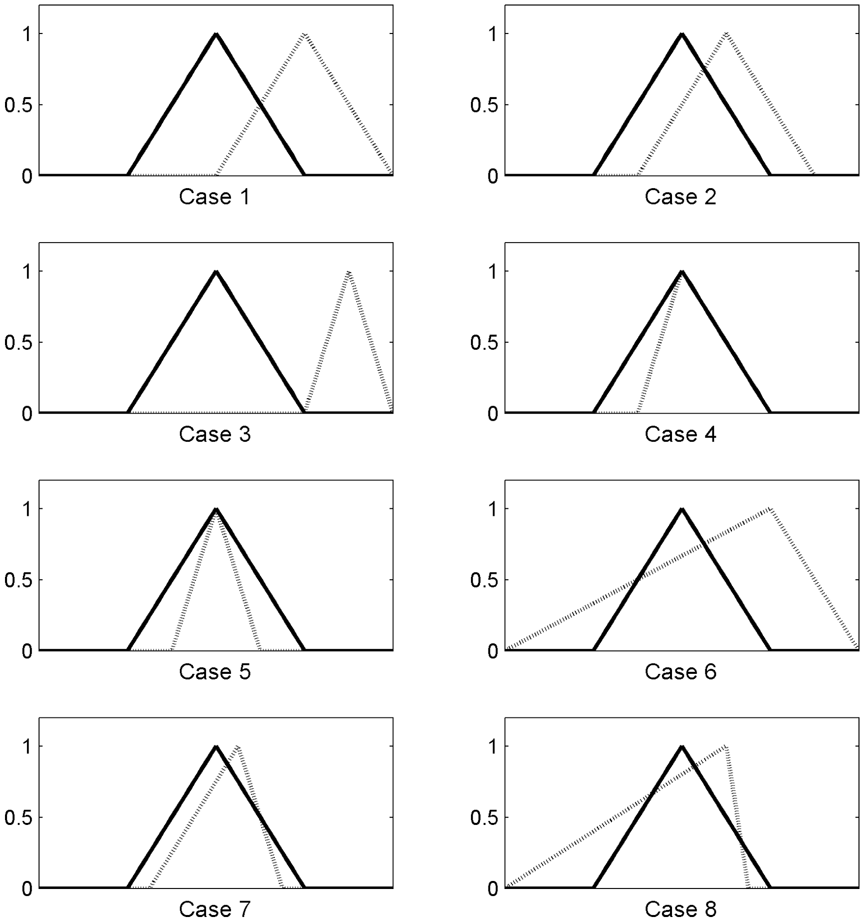



**Figure 30:** Case studies for the evaluation of the ordering of fuzzy sets $A_1$ (solid) and $A_2$ (dotted)

**Table 1:** Evaluation of $A_1 \leq A_2$ and $A_2 \leq A_1$ on the cases of Figure 30

| | Kolodziejczyk | | Yuan | | Jaccard |
|---|---|---|---|---|---|
| | $K_{\leq}(A_1, A_2)$ | $K_{\leq}(A_2, A_1)$ | $Y_{\leq}(A_1, A_2)$ | $Y_{\leq}(A_2, A_1)$ | $S_J(A_1, A_2)$ |
| 1 | 1 | 0 | 0.9080 | 0.0920 | 0.25 |
| 2 | 1 | 0 | 0.7444 | 0.2556 | 0.5625 |
| 3 | 1 | 0 | 1 | 0 | 0 |
| 4 | 1 | 0 | 0.5761 | 0.4239 | 1 |
| 5 | 0.5 | 0.5 | 0.5 | 0.5 | 1 |
| 6 | 0.8333 | 0.1667 | 0.6576 | 0.3424 | 0.4375 |
| 7 | 0.8333 | 0.1667 | 0.5737 | 0.4263 | 0.9375 |
| 8 | 0.4003 | 0.5997 | 0.4730 | 0.5270 | 0.6623 |

1. Methods which extract crisp quantities from fuzzy sets and then compare them (including the trivial approaches described above).

2. Methods which consider a third reference set.

3. Methods which evaluate the $\leq$ relation as a fuzzy relation (see Definition 28).

To our aims, methods in the third category are particularly attractive, since they provide us with enough expressibility to assess that $A_1 \leq A_2$ and $A_2 \leq A_1$ are true at different degrees, for instance 0.6 and 0.2, respectively. In this case, we write $R_{\leq}(A_1, A_2) = 0.6$ and $R_{\leq}(A_2, A_1) = 0.2$. If $R_{\leq}(A_1, A_2) + R_{\leq}(A_2, A_1) = 1$ holds, the relation is said to be *reciprocal*.

We remark that, whilst methods in the first and in the second categories can directly derive a total order among the fuzzy numbers in $\mathbb{R}$, methods in the third category need an extra defuzzification step to convert the fuzzy relation into a total order. Many properties have been defined for fuzzy ordering relations, but Wang proved that acyclicity is the only one that is required to derive a total order (Wan97).

All the approaches reviewed by Wang and Kerre were tested with respect to a set of reasonable axioms, and some methods from the first and third categories proved to be particularly effective.

We used the case studies of Figure 30 to benchmark the behavior of Kolodziejczyk's and Yuan's indices (Yua91), two reciprocal fuzzy ordering relations previously reviewed in (WK01b). Table 1 shows the values of the two indices for the eight case studies. The value of the Jaccard's similarity index is also shown as a reference.

Results highlight that, between the two indices, Yuan's index is able to distinguish one case from another in a "fuzzier" fashion, i.e., providing more information than Kolodziejczyk's one on the actual perceived ordering of the two fuzzy numbers.

Indeed, we observe that in cases 1 and 2 the fuzzy sets satisfy both the properties of distinguishability and coverage, whilst case 3 is lacking coverage and case 4 is lacking distinguishability. Kolodziejczyk's index returns $K_{\leq}(A_1, A_2) = 1$ in all four cases and, therefore, it cannot be used to evaluate distinguishability and coverage.

In contrast, Yuan's index is able to properly discriminate the four cases, giving a crisp value of 1 only in case 3, in which $A_1 \cap A_2 = \emptyset$. In the other cases, different degrees of truth are obtained by the evaluation of $Y_{\leq}(A_1, A_2)$. Thus, Yuan's index can be reasonably chosen as a building block for the definition of an interpretability index based on a fuzzy ordering relation.

### 4.3.2 The Interpretability Index

As we have detailed in the previous Sections, when assessing interpretability of a fuzzy partition in CA, we have to take coverage, distinguishability, and ordering into account. To this aim, we introduce the following novel index based on fuzzy ordering relations.

**Definition 43 (Interpretability index $\Phi_Q$)** *Given a partition $P = \{A_1, ..., A_i, ..., A_N\}$ consisting of N fuzzy sets, let $d_{ji} = |j - i|$ be the semantic distance between $A_j$ and $A_i$. The interpretability index $\Phi_Q$ is defined as*

$$\Phi_Q(P) = \frac{\displaystyle\sum_{\substack{1 \leq i \leq N-1 \\ i < j \leq N}} \frac{1}{d_{ji}} \mu_Q^{d_{ji}} \left( Q_{\leq}(A_i, A_j) \right)}{\displaystyle\sum_{\substack{1 \leq i \leq N-1 \\ i < j \leq N}} \frac{1}{d_{ji}}}, \tag{4.6}$$

*where $Q$ is a fuzzy ordering index and $\mu_Q^{d_{ji}}(x)$, with $x = Q_{\leq}(A_i, A_j)$, are fuzzy sets defined on the universe $[0, 1]$ of the values of $Q$.*

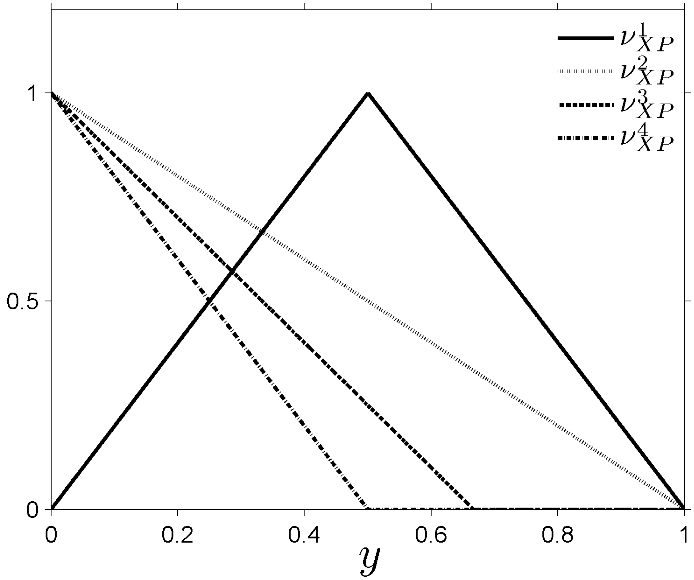
The *semantic distance* measures the linguistic proximity of terms. For instance, the semantic distance between $A_3$ and $A_1$ is 2. The value of $\Phi_Q(P)$ ranges between 0 (the lowest level of interpretability) and 1 (the highest level of interpretability). Thus, $\Phi_Q(P)$ should be close to 1 for uniform fuzzy partitions.

Fuzzy sets $\mu_Q^{d_{ji}}(x)$ are used to assess, for different values of $d_{ji}$, the value of the $Q$ index with respect to interpretability. For instance, in case of $d_{ji} = 1$ and $A_i \cap A_j = \emptyset$, i.e., a situation similar to case 3 in Figure 30, adjacent fuzzy sets are not overlapped and, therefore, coverage is not verified. It follows that $\mu_Q^{d_{ji}}(x)$ should return a value close to 0. On the other hand, in case of $d_{ji} > 1$ and $A_i \cap A_j = \emptyset$, to enforce distinguishability, the fuzzy sets should not be overlapped and, therefore, $\mu_Q^{d_{ji}}(x)$ should return a value close to 1. Obviously, the definition of the family of fuzzy sets $\mu_Q^{d_{ji}}(x)$ is a critical step, since these fuzzy sets represent the actual link between the evaluation of the ordering, coverage and distinguishability properties. In the following, we describe a procedure to generate the family of fuzzy sets $\mu_Q^{d_{ji}}(x)$.

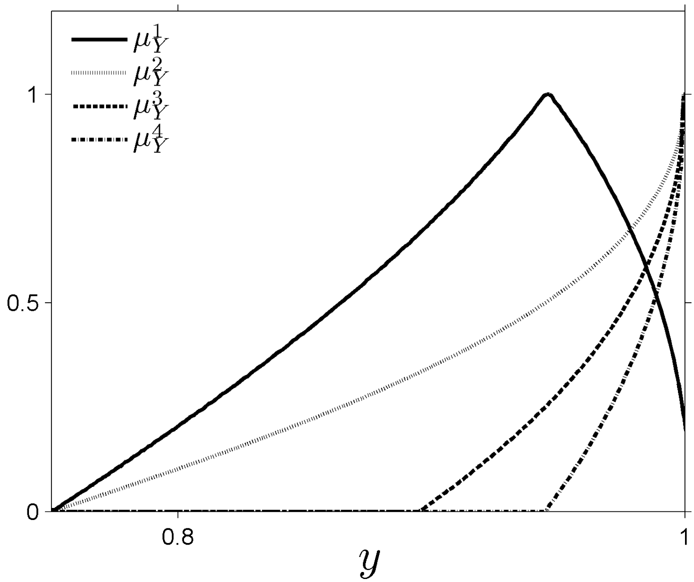
We start from the evaluation of the measure of interpretability introduced in Section 4.2, based on the value $y$ of the crossing point between two fuzzy sets (denoted as XP in the following). We can distinguish two cases.

- *Semantically adjacent fuzzy sets* ($d_{ji} = 1$). In this case, $y$ should be neither too close to 1 nor to 0, so as to preserve, respectively, the distinguishability and the coverage properties. On the other hand, the two properties are both verified when $y$ is close to 0.5. These observations can be modeled by the fuzzy set $\nu_{XP}^1(y)$ shown in solid line in Figure 31(a).

- *Semantically non adjacent fuzzy sets* ($d_{ji} > 1$). In this case, we do not care about coverage, since it is already ensured by adjacent fuzzy sets, but we stress distinguishability. Thus, $y$ should be close to 0. Nevertheless, depending on the actual value of $d_{ji}$, we can




(a) $\nu_{XP}^{d_{ji}}(y)$




(b) $\mu_Y^{d_{ji}}(x)$

**Figure 31:** The fuzzy sets $\nu_{XP}^{d_{ji}}(y)$ and $\mu_Y^{d_{ji}}(x)$ used to assess $\Phi_Y$

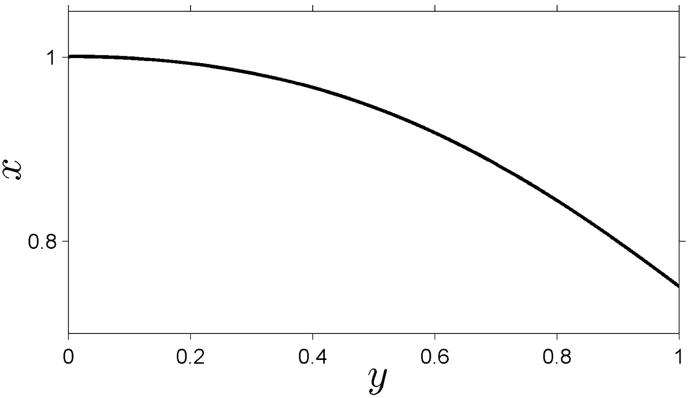



**Figure 32:** The empirical relation $R_{Y,XP}(x, y)$ used to assess $\Phi_Y$

still tolerate some overlapping between the two fuzzy sets, and this tolerance should decrease with the increase of $d_{ji}$. These observations can be modeled by the fuzzy set $\nu_{XP}^{d_{ji}}(y)$ shown in dotted line in Figure 31(a) for different values of $d_{ji}$. We note that, while $\nu_{XP}^{d_{ji}}(0) = 1 \; \forall d_{ji} > 1$, the right spread shrinks toward 0 as $d_{ji}$ increases, so as to reduce the tolerance.

To obtain $\mu_Q^{d_{ji}}(x)$, we project $\nu_{XP}^{d_{ji}}(y)$ from its original base variable $y$ to the base variable $x$. The overall process can be formalized as follows.

1. We choose $\nu_{XP}^{d_{ji}}(y)$ as triangular membership functions, defined by the three breakpoints $(sl^{d_{ji}}, c^{d_{ji}}, su^{d_{ji}})$. We set $sl^{d_{ji}} = 0 \; \forall d_{ji}$, $c^1 = 0.5$, $c^{d_{ji}} = 0 \; \forall d_{ji} > 1$, $su^1 = 1$, and $su^{d_{ji}} = 2/d_{ji} \; \forall d_{ji} > 1$. Figure 31(a) shows $\nu_{XP}^{d_{ji}}(y)$, with $N = 5$ and $d_{ji} = 1, ..., 4$.

2. We empirically identify a relation $R_{Q,XP}(x, y)$ by evaluating $x$ and $y$ for a number of differently overlapping trapezoidal membership functions.

3. Using the extension principle and $R_{Q,XP}(x, y)$, we project the fuzzy sets $\nu_{XP}^{d_{ji}}(y)$ from the base variable $y$ to the base variable $x$, thus obtaining a corresponding $\mu_Q^{d_{ji}}(x)$. More formally, we have

$$\mu_Q^{d_{ji}}(x) = \sup_{y \in [0,1]} \min(\nu_{XP}^{d_{ji}}(y), R_{Q,XP}(x,y)), \tag{4.7}$$

where $d_{ji} = 1, ..., N-1$.

We illustrate the above process by detailing the sets $\mu_Y^{d_{ji}}(x)$ corresponding to Yuan's fuzzy ordering index. We recall the formula to compute Yuan's index (Yua91; WK01b).

**Definition 44 (Yuan's ordering index)** *Given two fuzzy sets $A_1, A_2 \in \mathcal{F}(\mathcal{U})$, Yuan's ordering index $Y_{\leq}(A_1, A_2) : \mathcal{F}(\mathcal{U}) \times \mathcal{F}(\mathcal{U}) \to [0,1]$ is defined as*

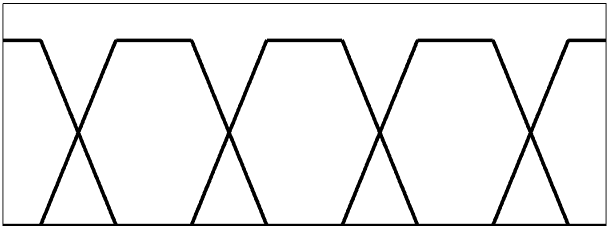
$$Y_{\leq}(A_1, A_2) = \frac{\Upsilon(A_2, A_1)}{\Upsilon(A_1, A_2) + \Upsilon(A_2, A_1)}, \tag{4.8}$$

*where $\Upsilon(A_1, A_2)$ is defined as*

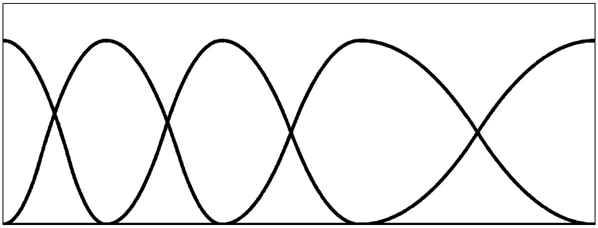
$$\Upsilon(A_1, A_2) = \int\limits_{\alpha | u_{A_{1\alpha}} > l_{A_{2\alpha}}} (u_{A_{1\alpha}} - l_{A_{2\alpha}}) d\alpha + \int\limits_{\alpha | l_{A_{1\alpha}} > u_{A_{2\alpha}}} (l_{A_{1\alpha}} - u_{A_{2\alpha}}) d\alpha,$$

*$\alpha \in [0,1]$ is an $\alpha$-cut value, $A_{i\alpha}$ is the crisp set obtained by $\alpha$-cutting $A_i$, $i = \{1,2\}$, and $l_{A_{i\alpha}}$ and $u_{A_{i\alpha}}$ are the lower and upper bounds of $A_{i\alpha}$, respectively.*

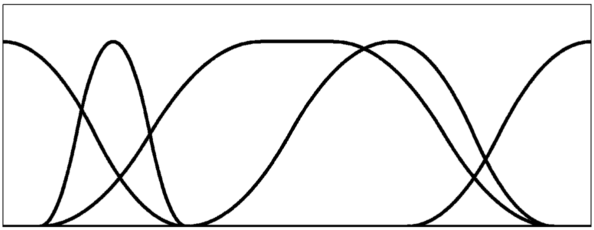
Figure 32 shows the empirical relation $R_{Y,XP}(x,y)$ used to project $\nu_{XP}^{d_{ji}}(y)$ on the base variable $x$. Figure 31(b) shows the projections $\mu_Y^{d_{ji}}(x)$ obtained by applying the overall process. This family of fuzzy sets can then be employed in the corresponding interpretability index $\Phi_Y(P)$. Finally, in Figure 33, we show the evaluation of the proposed index on four sample partitions characterized by different degrees of coverage, distinguishability and ordering of fuzzy sets.



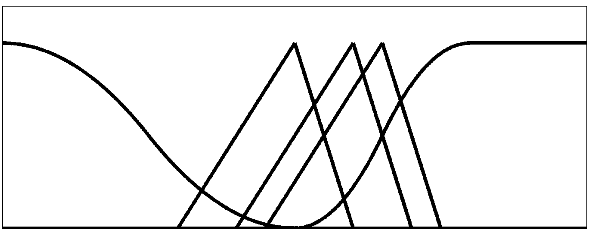
(a) $\Phi_Y = 1$



(b) $\Phi_Y = 0.9204$



(c) $\Phi_Y = 0.7166$



(d) $\Phi_Y = 0.4512$

**Figure 33:** Evaluation of $\Phi_Y$ on sample partitions with different degrees of coverage, distinguishability and ordering of fuzzy sets

# Chapter 5

# Evolutionary Algorithms for Automatic Context Adaptation of Fuzzy Systems

Thanks to their modeling capabilities, the scaling function $\psi$ and the four fuzzy modifiers introduced in Chapter 3 allow adapting normalized partitions to any context. Indeed, the combination of all the operators is able to generate a wide range of different context-adapted partitions. In real-world applications, we first have to determine whether each operator actually need to be applied to each fuzzy partition and, then, which parameter values have to be used for the selected ones. However, as stated in Section 2.1.4, FRBSs do not provide a native technique for such adaptive learning.

To the aim of generating context-adapted FRBSs, in this Chapter we introduce two evolutionary-based learning approches. In the framework of CA illustrated in Section 2.2, this process is classified as an instantiation process.

As typically performed in genetic learning of FRBSs, the above choices are based on the maximization of the accuracy of the GFRBS on real-world examples which are a sample of the effects of the context on the normal operation of the real system (CGH+04; Ish07; Her08). Furthermore, as discussed in Chapter 4, the partition generated by the application of the operators should satisfy the ordering and interpretability constraints.

Let us consider a data set $\mathcal{D} = \{(\hat{\mathbf{x}}_1, \hat{y}_1), ..., (\hat{\mathbf{x}}_d, \hat{y}_d), ..., (\hat{\mathbf{x}}_D, \hat{y}_D)\}$ of $D = |\mathcal{D}|$ real-world samples, where $\hat{\mathbf{x}}_d = (\hat{x}_{d1}, ..., \hat{x}_{dv}, ..., \hat{x}_{d(V-1)})$, $\hat{\mathbf{x}}_d \in \mathcal{U}_1 \times ... \times \mathcal{U}_v \times ... \times \mathcal{U}_{(V-1)}$, is a vector of $V-1$ input values, and $\hat{y}_d \in \mathcal{U}_V$ is the corresponding output. Then, the objective of our genetic learning process is to find a multi-input single-output GFRBS $F$ which correctly models $\mathcal{D}$ and, at the same time, takes the accuracy-interpretability tradeoff into account.

More formally, let us characterize $F$ by its input-output function

$$f_F(\mathbf{x}) : \mathcal{U}_1 \times ... \times \mathcal{U}_{(V-1)} \rightarrow \mathcal{U}_V, \tag{5.1}$$

and by the value of an index $I(F)$ which measures the overall interpretability of the GFRBS. It follows that the genetic learning process can be modeled as a *constrained optimization problem*, as

$$\begin{cases} \min \sum_{d=1}^{D} E\left(f_F\left(\hat{\mathbf{x}}_d\right), \hat{y}_d\right) \\ \iota_{\min} \leq I(F) \leq \iota_{\max}, \end{cases} \tag{5.2}$$

where $E$ is an error measure between the model output $f_F\left(\hat{\mathbf{x}}_d\right)$ and the corresponding real-world sample $\hat{y}_d$, and $[\iota_{\min}, \iota_{\max}]$ is the allowed range of values for $I(F)$. Conversely, if a special focus on interpretability is desired, the genetic learning process can be expressed in terms of a *multi-objective optimization problem*, as

$$\begin{cases} \min \sum_{d=1}^{D} E\left(f_F\left(\hat{\mathbf{x}}_d\right), \hat{y}_d\right) \\ \max I(F). \end{cases} \tag{5.3}$$

From an EA point-of-view, the alternative formulations of the learning process reflect into the choice between two evolutionary strategies.

- *Constrained single-objective genetic algorithm.* As regards Equation 5.2, a well-known solution for single-objective constrained optimization via EAs is the use of *penalty functions* (MS96; Coe02). Hence, in this approach, while searching for context-adapted GFRBSs which minimize the performance error, we measure the interpretability level of the linguistic variables. Then, we penalize individuals that show a low level of interpretability. This approach is detailed in Section 5.1.
- *Multi-objective evolutionary algorithm.* As regards Equation 5.3, we can perform a multi-objective search, so as to concurrently reduce the error between the system and the expected outputs and increase the level of interpretability. Actually, this corresponds to the identification of a set of (approximated) *Pareto-optimal* GFRBSs that is commonly achieved by a MOEA (DAPM02; IN07). This approach is described in Section 5.2.

Both the alternatives for instantiation can be described by the common framework depicted in Figure 34. Indeed, we always start from a normalized FRBS, i.e., an FRBS which adopts linguistic terms associated with fuzzy sets uniformly distributed on a normalized universe of discourse $\mathcal{N} = [0,1]$. As largely discussed in Section 2.2, the normalized FRBS can be provided by an expert (via knowledge elicitation) or computed by using some automatic identification method (via abstraction).

The normalized FRBS is input to the *Context Adaptation Module*. By using experimental data collected by letting the real system operate in the context to be modeled (i.e., $\mathcal{D}$), the *Optimization Module* searches for the best parameters to be used in the operators so as to adapt the normalized FRBS to the given context. In this work, the *Optimization Module* is implemented by one of the two evolutionary approaches detailed in the rest of this Chapter. Indeed, the individuals comprised in the populations of the EAs represent the values of the parameters which define the operators used to adapt Mamdani-type GFRBSs.

Finally, the output of the *Context Adaptation Module* is a context-adapted FRBS, that is, an FRBS with the same RB as the normalized one, but with the DB adapted to the given context.

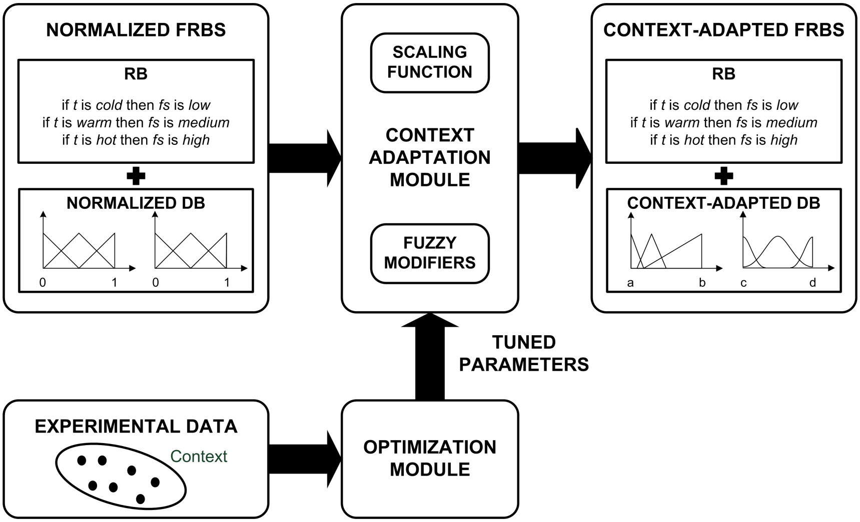



**Figure 34:** The overall instantiation process based on evolutionary algorithms

# 5.1 Constrained Single-Objective Genetic Algorithm

In this Section, we discuss the development of a GFRBS based on constrained single-objective genetic algorithm (SOGA) for context adaptation. Preliminary versions of the proposed technique can be found in (BLM06a; BLM06b; BLM08).

## 5.1.1 Chromosome Encoding

The chromosome which represents a context-adapted DB is composed of $V$ strings of 77 bits, where each string encodes five control genes and nine 8-bit parameters. The five control genes are used to select which operators have to be applied on the $v$-th partition, while the other 72 bits are used to encode the values of the parameters assigned to the tuning operators.

The first $V-1$ strings and the last string determine the parameters for the input variables and for the output variable, respectively.

As shown in Figure 35, each string is characterized by a hierarchical structure: the first five bits, one for each tuning operator, control whether the corresponding operator is applied or not on each fuzzy partition. The other 72 bits are organized in sub-strings of eight bits. Each sub-string determines the value of a given parameter. The chromosome has the following structure

$$\begin{aligned}
&(C_{SF_1}, C_{CP_1}, C_{CW_1}, C_{SW_1}, C_{GP_1},\\
&b(u_{\min_1}), b(u_{\max_1}), b(\lambda_1), b(k_{SF_1}),\\
&b(k_{CP_1}), b(k_{CW_1}), b(k_{SW_1}), b(\theta_1), b(k_{GP_1}),\\
&\ldots\\
&C_{SF_V}, C_{CP_V}, C_{CW_V}, C_{SW_V}, C_{GP_V},\\
&b(u_{\min_V}), b(u_{\max_V}), b(\lambda_V), b(k_{SF_V}),\\
&b(k_{CP_V}), b(k_{CW_V}), b(k_{SW_V}), b(\theta_V), b(k_{GP_V})),
\end{aligned} \tag{5.4}$$

where $C_{SF_v}$, $C_{CP_v}$, $C_{CW_v}$, $C_{SW_v}$, and $C_{GP_v}$, with $v = 1, ..., V$, are the control genes that determine whether, respectively, the non linear scaling function $\psi$, and the core-position, the core-width, the support-width, and the generalized positively modifiers have to be applied to the $v$-th partition. The $b(u_{\min_v})$, ..., $b(k_{SF_v})$ and the $b(k_{CP_v})$, ..., $b(k_{GP_v})$ are, respectively, the binary encodings of the values of the parameters of $\psi$ and of the four fuzzy modifiers for the $v$-th partition.

Instead of using a mixed binary-real variable coding (with five binary genes for the choice of the tuning operators and nine real genes for the parameter values), we decided to adopt a binary-coded GA based on the following considerations. We observed that the quantization performed by binary coding does not affect the precision of the choice of the parameter values due to the small ranges of these parameters. Furthermore, binary coding provides a discretization of the search space, thus allowing the exploration of the solution space with lower computational effort.

However, binary coding suffers from the following problem: mating can generate descendants which inherit no characteristics of the parents.

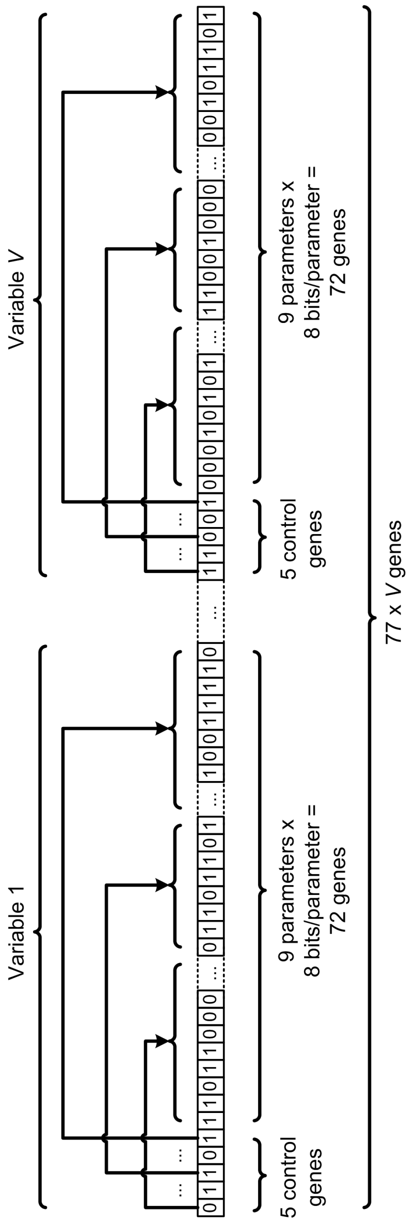



**Figure 35:** The structure of the chromosome

**Table 2:** The structure of the $v$-th string of the chromosome

| Operator | Parameter | Range |
|---|---|---|
| *Scaling function* | $u_{\min_v}$ | $[\min_d(\hat{x}_{dv}) - \delta\Delta_v, \min_d(\hat{x}_{dv}) + \delta\Delta_v]$ |
| | $u_{\max_v}$ | $[\max_d(\hat{x}_{dv}) - \delta\Delta_v, \max_d(\hat{x}_{dv}) + \delta\Delta_v]$ |
| | $\lambda_v$ | $[0, 1]$ |
| | $k_{SF_v}$ | $[0.4, 2]$ |
| *CP modifier* | $k_{CP_v}$ | $[-0.9, 0.9]$ |
| *CW modifier* | $k_{CW_v}$ | $[-1, 0.25]$ |
| *SW modifier* | $k_{SW_v}$ | $[0.667, 2]$ |
| *GP modifier* | $\theta_v$ | $[0, 1]$ |
| | $k_{GP_v}$ | $[0.75, 4]$ |

To solve this problem, we have adopted, as usual when using binary chromosomes, the Gray decoding to generate individuals from chromosomes (Mic99).

Table 2 shows the lower and upper bounds of the range of possible values for each parameter coded in the corresponding sub-string. We observe that these ranges are actually sub-intervals of the domains of definition of the parameters: the lower and upper bounds of the ranges have been chosen heuristically so as to avoid values of parameters which make the final partition non interpretable.

In Table 2, $\Delta_v$ denotes the difference between the maximum and the minimum values of the $v$-th variable in $\mathcal{D}$, i.e.

$$\Delta_v = \max_{d=1,\dots,D}(\hat{x}_{dv}) - \min_{d=1,\dots,D}(\hat{x}_{dv}), \ \forall v = 1, ..., V-1, \tag{5.5}$$

and similarly for $v = V$ (with $\hat{y}_d$ in place of $\hat{x}_{dv}$). Further, $\delta \in [0, 0.5]$ is a design parameter. In the experiments, we set $\delta$ to 0.15.

Note that some values of the $v$-th variable in $\mathcal{D}$ might be outside of the interval $[u_{\min_v}, u_{\max_v}]$: actually, these values are probably outliers. Thus, the appropriate choice of $\delta$ avoids that possible outliers affect the context modeling capabilities of our approach. Anyway, at runtime, if a value of an input variable is smaller (higher) than the lower (upper) bound of $\mathcal{U}_v$, it is assigned to the first (last) MF with membership value equal to 1.

### 5.1.2 Phenotype Decoding

The value $k$ of a generic parameter is determined by the following formula

$$k = k_{\min} + (k_{\max} - k_{\min}) \frac{g\left(b\left(k\right)\right)}{2^8 - 1}, \tag{5.6}$$

where [$k_{\min}$,$k_{\max}$] is the interval of values of parameter $k$ as defined in Table 2, $b(k)$ is the 8-bit string representing $k$ in the chromosome, and $g$ is a function that maps an 8-bit sub-string into its Gray integer value.

The context-adapted partitions are instantiated by applying the five operators to the normalized FRBS, as depicted in Figure 34. In generating the context-adapted partitions, we first apply the scaling function, and then the four modifiers, whose order of application is not significant (see Section 3.2.5).

We remark that, in order to cover the overall universe of discourse, even when the control gene of the non linear scaling function is off (i.e., $C_{SF_v} = 0$), we always perform a linear scaling from the normalized interval $[0, 1]$ to the interval $[u_{\min_v}, u_{\max_v}]$ encoded in the chromosome. We remark that this corresponds to applying the scaling function $\varphi_0$ introduced in (GG94). Conversely, if the control gene is on, we apply the non linear scaling with the value of parameters $\lambda_v$ and $k_{SF_v}$ determined by Equation 5.6.

Some problems about the normality of the fuzzy partition may arise when both the core-position and the core-width modifiers are applied, because these modifiers may move the core of the first and/or the last fuzzy set out of the bounds of the universe of discourse, thus leaving some subnormal fuzzy sets in the partition. To avoid this problem, once we have applied the modifiers, we adjust the bounds of the universe of discourse so as to include the upper and lower bounds of the cores of the first and last fuzzy sets, respectively.

### 5.1.3 Fitness Function

Since we adopt a constrained GA based on penalty functions, the fitness function is defined as

$$f_{GA}(F, \mathcal{D}) = \sum_{d=1}^{D} E\left(f_F\left(\hat{\mathbf{x}}_d\right), \hat{y}_d\right) + P_{GA}(F), \tag{5.7}$$

where $P_{GA}$ is a properly defined penalty function employed to enforce the interpretability constraints. To this aim, we compute the index $\Phi_{XP}$ defined in Section 4.2 for all the partitions of $F$.

As regards the error measure, we observe that it is strongly application-dependent. For instance, in control applications, the *integral of time and absolute error* (ITAE) is commonly used (Kar91; Mag02). In regression and modeling applications, a widely employed accuracy measure is the *mean square error* (MSE), defined as

$$MSE(F, \mathcal{D}) = \frac{1}{2D} \sum_{d=1}^{D} \left(f_F\left(\hat{\mathbf{x}}_d\right) - \hat{y}_d\right)^2. \tag{5.8}$$

Hence, the overall formulation of our fitness function is

$$f_{GA}(F, \mathcal{D}) = \frac{1}{2D} \sum_{d=1}^{D} \left(f_F\left(\hat{\mathbf{x}}_d\right) - \hat{y}_d\right)^2 + \beta \sum_{v=1}^{V} \Phi_{XP}(P_v), \tag{5.9}$$

where $P_v$ is the fuzzy partition of the linguistic variable bound to the $v$-th variable of the GFRBS $F$, and $\beta \geq 0$ is a tunable design parameter which controls the influence of the penalties with respect to the accuracy measure.

### 5.1.4 Genetic Evolution

At generation $t = 0$, we start with an initial population $P(0)$ composed of $N_{\text{pop}}$ individuals generated by a random uniform distribution. At each generation, the simple *uniform crossover* and the *uniform mutation* operators are applied, with a crossover and a mutation probability of, respectively, 0.8 and 0.05 (see (Mic99) for a review of genetic operators).

Chromosomes to be mated are chosen by using the well-known *deterministic tournament selection* method, with a tournament fraction of 1/8 of the population.

Further, we adopt the following acceptance mechanism: the new population $P(t+1)$ is composed of offspring, except for a percentage of 5% of the best individuals of population $P(t)$.

When the average of the fitness values of all the individuals in the population is greater than 99.9% of the fitness value of the best individual or a prefixed number of $T_{\max}$ generations has passed, the GA is considered to have converged.

## 5.2 Multi-Objective Evolutionary Algorithm

In this Section, we approach the instantiation of the Mamdani-type GFRBS as the multi-objective optimization problem shown in Equation 5.3. This technique has been firstly introduced in (BLMS07a) and later exploited in (BLMS08; BDLM08).

### 5.2.1 The Non-Dominated Sorting Genetic Algoritm II

To perform the multi-objective search, we adopt a MOEA, namely the NSGA-II developed by Deb et al. in (DAPM02).

NSGA-II is a fast and elitist EA that evolves a population of possible solutions to multi-objective problems. The rank assignment of the algorithm is based on the concept of *Pareto-dominance*: a solution $s_1$ dominates a solution $s_2$ if and only if $s_1$ is worse than $s_2$ in no objective and $s_1$ is better than $s_2$ in at least one objective. The set of Pareto-optimal solutions (i.e., a set of solutions which are not dominated by any another one) is called *Pareto front*. Furthermore, to allow a balanced exploration of the fronts, NSGA-II exploits an ad-hoc density-estimation metric, called *crowding distance*.

NSGA-II starts from an initial random population $P(0)$ of $N_{\text{pop}}$ individuals sorted according to the non-dominance. Each solution is associated with a rank equal to its non-dominance level (e.g., rank 1 is given to solutions which are non-dominated, rank 2 is given to the solutions dominated by just another one, and so on).

At each generation $t, t = 0, ..., T_{\max}$, an offspring population $Q(t)$ of size $N_{\text{pop}}$ is produced by selecting mating individuals through the *binary tournament selection*, and by applying the crossover and mutation operators. The parent population $P(t)$ and the offspring population $Q(t)$ are combined so as to generate a new population $R(t) = P(t) \cup Q(t)$. Then, a rank is assigned to each individual in $R(t)$. Based on these ranks, $R(t)$ is split into different non-dominated fronts, one for each different rank. Within each front, a specific crowding measure, which represents the sum of the distances to the closest individual along each objective, is used to define an ordering among individuals.

The new parent population $P(t+1)$ is generated by deleting from $R(t)$ the worst $N_{\text{pop}}$ individuals (considering first the ordering among the fronts and then that among the individuals). The algorithm terminates when it reaches the maximum number $T_{\max}$ of generations.

### 5.2.2 Context Adaptation through NSGA-II

In the last years, NSGA-II has been successfully applied to a number of GFRBS (IN07). Hence, we exploited it for our MOEA-based CA approach.

Basically, the chromosome, the phenotype decoding, and the crossover and mutation operators are the same of the single-objective version, as explained in Sections 5.1.1, 5.1.2, and 5.1.4 respectively.

On the other hand, the two objectives of NSGA-II are the minimization of the $MSE$ shown in Equation 5.8 and the maximization of the averaged value of the index $\Phi_Y$ (see Section 4.3) computed for all input and output partitions of the FRBS, as

$$\bar{\Phi}_Y(F) = \frac{1}{V} \sum_{v=1}^{V} \Phi_Y(P_v). \tag{5.10}$$

Again, we start with an initial population composed of randomly generated individuals. In this GRBS, selection is performed by the standard binary tournament proposed by (DAPM02) in the original version of NSGA-II.

# Chapter 6

# Experimental Results

In the previous Chapters, we have reviewed existing CA techniques and introduced new tools, indices, and learning approaches to the aim of augmenting the instantiation process of GFRBSs.

In the following, we assess the effectiveness of our overall proposal by applying it to four context-aware data sets. Moreover, we provide an analysis of the effects of each CA operator and detailed benchmarks with some of the existing approaches.

## 6.1 Data Sets

In this Section, we introduce the context-aware data sets that we will exploit in the remaining of the Chapter for numerical evaluation of CA approaches. We remark that finding a proper data set to assess CA algorithms is not an easy task, given the very specific kind of application and since there not exist relevant benchmarks in the literature. Indeed, the data sets have to be organized in different contexts and, further, the data collected in different contexts must all resemble a similar instance of the same system.

In the following, we will describe four data sets which enforces the above requirements. The first application is the only context-aware benchmark existing in the literature that can be used to our aims. This data

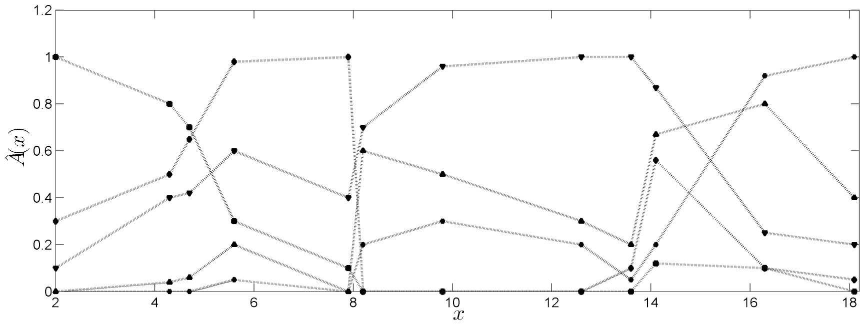



**Figure 36:** Plot of the context-adapted fuzzy partition data set

set associates points of a universe of discourse to values of arbitrarily context-adapted fuzzy sets. Although simple, it can be exploited to compare CA approaches on a single partition, without defining an RB and, thus, an FRBS.

The remaining three data sets are a two-input regression problem, and two modeling applications: a simple relation between years of experience and wage contextualized into four different education levels of workers, and a challenging four-input modeling problem regarding fuel consumption in highway and city.

### 6.1.1 Context-Adapted Fuzzy Partition

This data set was introduced in (PGG97) and consists of elements situated in a segment $[2, 18.1] \in \mathbb{R}$ that are assigned to five fuzzy sets. In a sense, such data set can be interpreted as a collection of twelve sampling points taken from a context-adapted fuzzy partition. The data set is shown in Table 3 and plotted in Figure 36.

We remark that the algorithms developed in Chapter 5 cannot be directly applied to this data set. Indeed, differently from the ones introduced in the following Sections, this application does not represent a set $\mathcal{D}$ of input-output pairs.

**Table 3:** The context-adapted fuzzy partition data set

| $x$ | $\hat{A}_1(x)$ | $\hat{A}_2(x)$ | $\hat{A}_3(x)$ | $\hat{A}_4(x)$ | $\hat{A}_5(x)$ |
|---|---|---|---|---|---|
| 2.0 | 1 | 0.3 | 0.1 | 0 | 0 |
| 4.3 | 0.8 | 0.5 | 0.4 | 0.04 | 0 |
| 4.7 | 0.7 | 0.65 | 0.42 | 0.06 | 0 |
| 5.6 | 0.3 | 0.98 | 0.6 | 0.2 | 0.05 |
| 7.9 | 0.1 | 1 | 0.4 | 0 | 0 |
| 8.2 | 0 | 0 | 0.7 | 0.6 | 0.2 |
| 9.8 | 0 | 0 | 0.96 | 0.5 | 0.3 |
| 12.6 | 0 | 0 | 1 | 0.3 | 0.2 |
| 13.6 | 0 | 0.1 | 1 | 0.2 | 0.05 |
| 14.1 | 0.12 | 0.56 | 0.87 | 0.67 | 0.2 |
| 16.3 | 0.1 | 0.1 | 0.25 | 0.8 | 0.92 |
| 18.1 | 0 | 0.05 | 0.2 | 0.4 | 1 |

### 6.1.2 The Structure of Wages

In economics, The structure of wages is represented by families of curves which show how the hourly wage of workers (HW) changes with the amount of years of experience (YE). In this kind of data, context is well represented by some common features shared among the people being interviewed (e.g., age, sex, social conditions, etc).

The four curves shown in Figure 37 represent the structure of wages for *college graduates*, *college drop-outs*, *high-school graduates* and *high-school drop-outs*. Hence, we consider the educational level as the context. This example is inspired by (MW92): we chose 100 points for each context and added to the curves some Gaussian noise with zero mean and variance equal to 1% of the range of HW-values of each curve. Further, the data set was split into five folds to perform cross-validation. Hence, each fold is composed of 80 training points and 20 test points.

By observing the common trend of the curves, we can derive a set of linguistic rules which reproduce the knowledge of an expert. Table 4 shows the chosen set, that includes six linguistic terms in YE (namely, *low*,

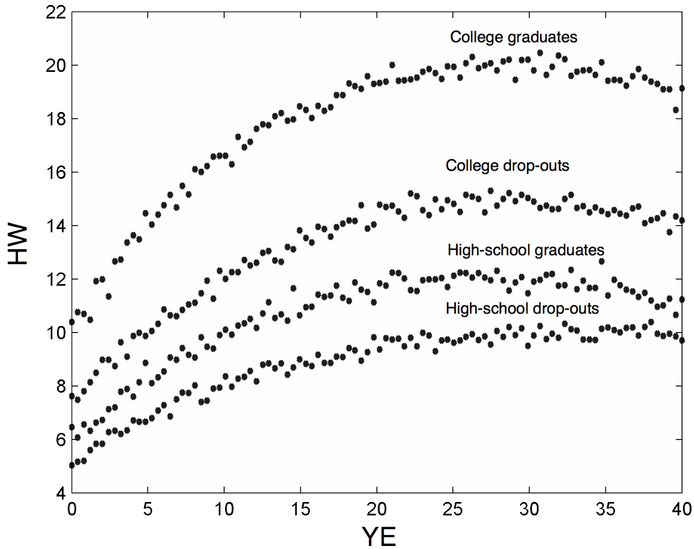



**Figure 37:** Plot of the structure of wages data set

*low-medium*, *medium-low*, *medium-high*, *high-medium*, and *high*) and five in HW (namely, *low*, *medium-low*, *medium*, *medium-high*, and *high*). The rules are all in the form "if YE is ... then HW is ...".

The normalized FRBS that we employ as universal model in our experiments is a single-input single-output Mamdani-type FRBS, with the RB as in Table 4 and the DB composed of two fuzzy partitions associated with, respectively, the input variable YE and the output variable HW.

### 6.1.3 Parametric Function

Let us consider the following parametric function:

$$g(x_1, x_2) = \kappa + e^{-(\kappa x_1)^{2\kappa} - (1+x_2)^{2\kappa}} - e^{-x_1^{2\kappa} - x_2^{2\kappa}} - e^{-(1+x_1)^{2\kappa} - (\kappa x_2)^{2\kappa}}, \quad (6.1)$$

where $x_1, x_2 \in [-1.5, 0.5]$. We generated three different instances of the curve by varying $\kappa$ in $\{2, 5, 7\}$. Figure 38 shows the isolevel contours produced by these instances.

Table 4: RB for the structure of wages data set

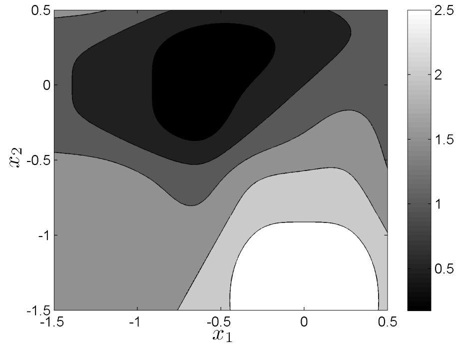
| Rule | YE | HW |
|---|---|---|
| $R_1$ | low | low |
| $R_2$ | low-medium | medium-low |
| $R_3$ | medium-low | medium |
| $R_4$ | medium-high | medium-high |
| $R_5$ | high-medium | high |
| $R_6$ | high | medium-high |

We can observe that, albeit the range of the output variable and the smoothness of each curve are affected by the value of parameter $\kappa$, all of the instances show a similar shape. In other words, the three curves may be considered as different context-adapted instances of the same system. In particular, different contexts are bound to different values of $\kappa$. Hence, we can define a set of rules that linguistically describe this common behavior, and regard each of the three curves as a different instantiation of the same generic shape in a different context, determined by the given value of $\kappa$.

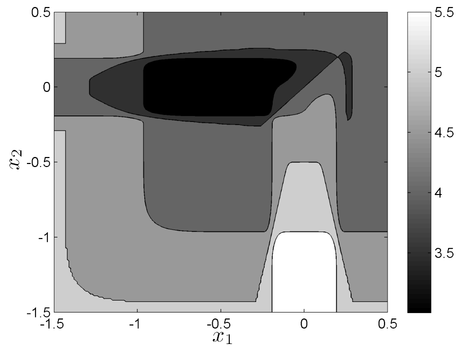
Table 5 shows the RB which describes the $\kappa$-independent behavior of $g$. In the RB, four linguistic labels are used in each universe, namely *low*, *medium-low*, *medium-high* and *high*. The rules are all in the form "if $x_1$ is ... and $x_2$ is ... then $y$ is ...".

Again, the universal model is represented by a Mamdani-type FRBS, with two input variables and one output variable. Since there are four linguistic labels for each variable in the RB, the normalized universes of discourse are uniformly partitioned into four trapezoidal fuzzy sets.

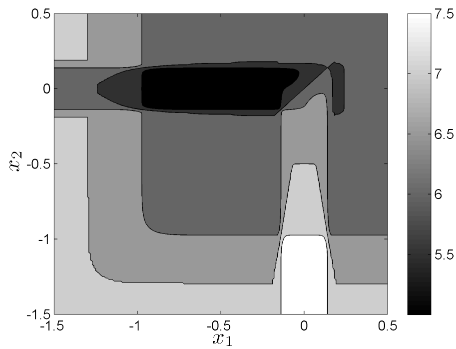
The data set was generated by assessing $g$ in a grid of 120 equally spaced points chosen in the $[-1.5, 0.5] \times [-1.5, 0.5]$ region. To avoid biasing the numerical evaluations with respect to a single training set, we divided the original data set into five subsets of 24 points each. Then, we generated a cross-validation data set composed of five folds, each with 96 and 24 training and test points, respectively.




(a) $\kappa = 2$




(b) $\kappa = 5$




(c) $\kappa = 7$

**Figure 38:** Isolevel contours of the parametric function data set

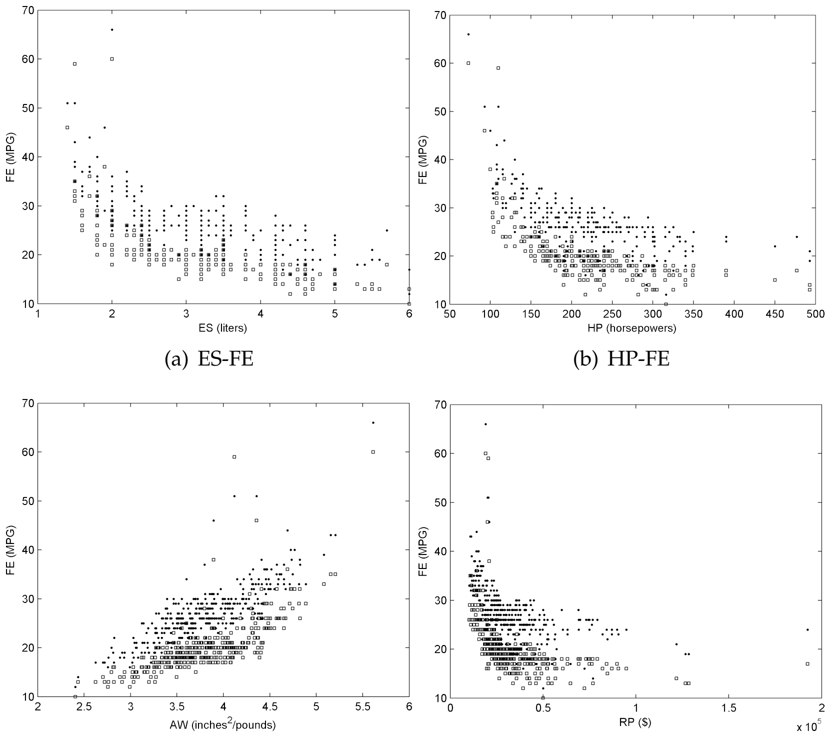
**Table 5:** RB for the parametric function data set

| **Rule** | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| $R_1$ | low | low | medium-high |
| $R_2$ | low | medium-low | medium-high |
| $R_3$ | low | medium-high | medium-low |
| $R_4$ | low | high | medium-low |
| $R_5$ | medium-low | low | medium-high |
| $R_6$ | medium-low | medium-low | medium-low |
| $R_7$ | medium-low | medium-high | low |
| $R_8$ | medium-low | high | low |
| $R_9$ | medium-high | low | high |
| $R_{10}$ | medium-high | medium-low | medium-high |
| $R_{11}$ | medium-high | medium-high | medium-low |
| $R_{12}$ | medium-high | high | low |
| $R_{13}$ | high | low | high |
| $R_{14}$ | high | medium-low | medium-high |
| $R_{15}$ | high | medium-high | medium-low |
| $R_{16}$ | high | high | medium-low |

### 6.1.4 Fuel Consumption

The 2004 new car and truck data set (Joh04) contains the features of a set of 428 different models of cars and trucks, such as *engine size* (ES), *horsepower* (HP), *retail price* (RP) and *fuel efficiency* (FE), in city and highway traffic conditions.

We preprocessed the data set by selecting the 387 vehicles with the complete set of 19 features. We aimed to model the effects of the traffic conditions on FE with respect to the other features. By observing such correlations, we realized that only ES, HP, RP and the ratio AW between base area and weight are strongly correlated to FE. Actually, AW was purposely generated by combining three features, namely width, length and weight. Hence, we decided to use only ES, HP, AW and RP as input variables and FE as output variable of our model. The city and highway traffic conditions were considered as two different contexts.



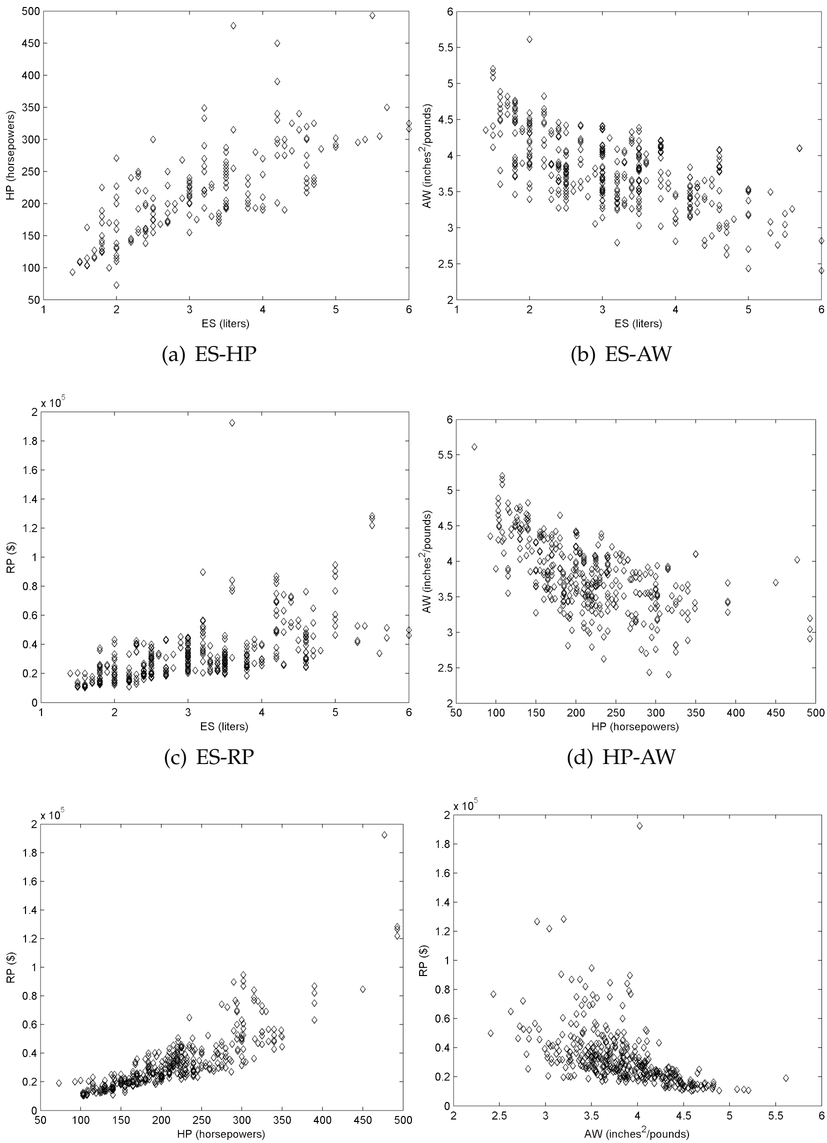

(a) ES-FE (b) HP-FE

(c) AW-FE (d) RP-FE

**Figure 39:** Plot of the fuel efficiency data set in the (a) ES-FE, (b) HP-FE, (c) AW-FE and (d) RP-FE planes

Figure 39 shows the distribution of the car and truck models in the ES-FE, HP-FE, AW-FE and RP-FE planes, respectively. ES, HP, AW, RP and FE are measured in liters, horsepower, inches$^2$/pounds, dollars and miles per gallon (MPG), respectively. The squares and the circles represent FE in, respectively, city and highway traffic conditions. The plots clearly show the existing correlation between the input variables and FE.

Figure 40 show the distribution of the car and truck models in the ES-HP, ES-AW, ES-RP, HP-AW, HP-RP and AW-RP planes, respectively. Note that, even though the relations between the features are evident, the data points are quite disperse in all the six planes.




(a) ES-HP (b) ES-AW

(c) ES-RP (d) HP-AW

(e) HP-RP (f) AW-RP

**Figure 40:** Plot of the fuel efficiency data set in the (a) ES-HP, (b) ES-AW, (c) ES-RP, (d) HP-AW, (e) HP-RP and (f) AW-RP planes

**Table 6:** RB for the fuel efficiency data set

| Rule | ES | HP | AW | RP | FE |
|---|---|---|---|---|---|
| $R_1$ | low | low | medium | low | high |
| $R_2$ | low | low | high | low | high |
| $R_3$ | low | medium | medium | low | medium |
| $R_4$ | low | medium | high | low | high |
| $R_5$ | medium | medium | low | medium | medium |
| $R_6$ | medium | medium | low | high | medium |
| $R_7$ | medium | medium | medium | medium | medium |
| $R_8$ | medium | medium | high | medium | medium |
| $R_9$ | medium | high | low | medium | medium |
| $R_{10}$ | medium | high | low | high | low |
| $R_{11}$ | medium | high | medium | medium | medium |
| $R_{12}$ | medium | high | medium | high | low |
| $R_{13}$ | medium | high | high | medium | medium |
| $R_{14}$ | high | medium | low | medium | medium |
| $R_{15}$ | high | medium | low | high | low |
| $R_{16}$ | high | medium | medium | medium | medium |
| $R_{17}$ | high | medium | medium | high | low |
| $R_{18}$ | high | high | low | medium | low |
| $R_{19}$ | high | high | low | high | low |
| $R_{20}$ | high | high | medium | medium | medium |
| $R_{21}$ | high | high | medium | high | low |

The rules of the normalized Mamdani-type FRBS, shown in Table 6, were extracted from the following intuitive considerations derived from experience, as

> FE decreases with the increase of ES, HP and RP, and increases with the increase of AW.

Further, we did not generate rules for meaningless or incompatible cases, such as *high* ES and *low* HP, or *low* ES and *high* RP.

We uniformly partitioned the normalized input variables and the output variable into three fuzzy sets, namely *low*, *medium*, and *high*. The number of fuzzy sets was chosen by interviewing a pool of experts and asking them for a meaningful partition of the universes. The rules are in the form "if ES is ... and HP is ... and AW is ... and RP is ... then FE is ...".

As in the previous data sets, the training and the test sets were obtained by applying a 5-fold cross-validation strategy.

## 6.2 Numerical Evaluations

In the following, we show extended results of the application of the two alternative instantiation techniques defined in Chapter 5 to the experimental data sets presented in Section 6.1. In the following, we refer to the single- and to the multi-objective versions of the algorithm as SOGA $\psi + \mathrm{FM}$ and MOEA $\psi + \mathrm{FM}$, respectively.

The tests were performed by running software implementations of the GFRBSs. The software has been developed on the *Matlab R2006a* platform (Mat06), exploiting some library functions of the *Genetic Algorithms and Direct Search Toolbox* and of the *Fuzzy Logic Toolbox*. Further, some functions of the *NSGA-II Matlab implementation* (Sch04) were used in the MOEA-based version of the GFRBS.

To the aim of testing the robustness of our approaches with respect to changes of the genetic meta-parameters, the experiments were run with different values of $N_{\mathrm{pop}}$ and $T_{\max}$. Further, in some cases our approaches were benchmarked with the following CA techniques inspired by the literature.

1. The absolute limit context determination with linear scaling introduced in (GG94), denoted in the following as AL $\varphi_0$.
2. The non linear scaling function $\varphi_2$ proposed in (CHMV01), where the parameters $u_{\min}$, $u_{\max}$, $a$, and $S$ are optimized similarly to SOGA $\psi + \mathrm{FM}$, i.e. by a SOGA with binary representation of parameters in the chromosome, uniform mutation, uniform crossover and binary tournament selection. We refer to this approach as SOGA $\varphi_2$.

As regards $\varphi_2$, we remark that we use just the scaling function and not the overall methodology introduced in (CHMV01) which, unlike our approach, does not rely on a universally valid RB, but rather identifies rules by exploiting a quick ad-hoc learn-by-example method similar to (WM92). Further, we remark that, to guarantee a fair comparison, we used for SOGA $\varphi_2$ the same chromosome coding, genetic operators, and parameter values that we applied in SOGA $\psi + \mathrm{FM}$.

### 6.2.1 Assessment of CA Operators

We exploited the context-adapted fuzzy partition data set to assess the contribution of each of the five operators introduced in Chapter 3 to the instantiation of a normalized fuzzy partition.

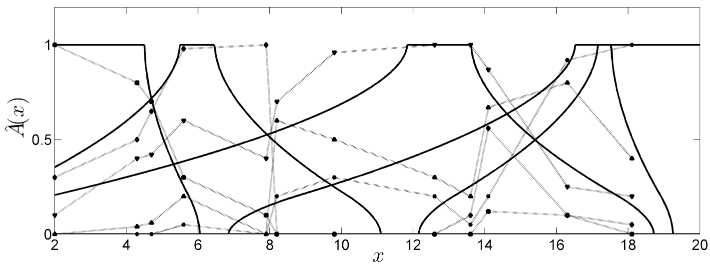
Since, as detailed in Section 6.1.1, the dataset comprises five linguistic terms, we started from a normalized uniform partition with five trapezoidal fuzzy sets, like the one depicted in Figure 12. Then, we applied the SOGA defined in Section 5.1 with different combinations of the five operators.

Obviously, in this application we do not adapt a whole FRBS but, rather, just a single partition. Hence, we cannot use the fitness function defined in Equation 5.9 and we have to define a different one.

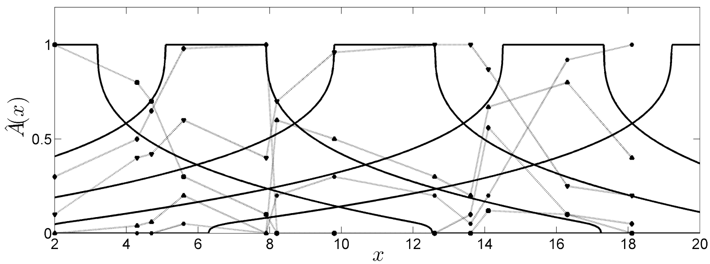
Let us denote as $\tilde{P}\{\tilde{A}_1, ..., \tilde{A}_5\}$ and $\hat{P}\{\hat{A}_1, ..., \hat{A}_5\}$ the context-adapted fuzzy partition resulting by the application of our SOGA and the original one sampled in the data set, respectively. The chosen fitness function is the total $MSE$, computed as

$$f_{GA}(\tilde{P}, \hat{P}) = \frac{1}{ND} \sum_{n=1}^{N} \sum_{d=1}^{D} \left( \tilde{A}_n\left(x_d\right) - \hat{A}_n\left(x_d\right) \right)^2, \qquad (6.2)$$

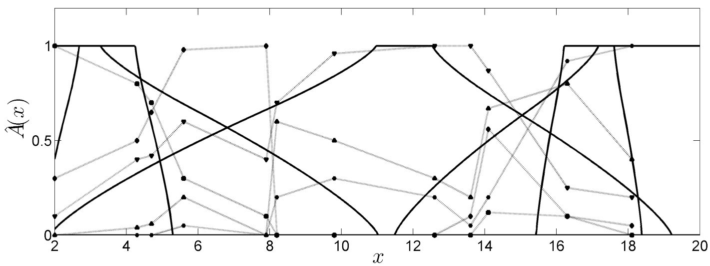
where, in this case, $N = 5$ and $D = 12$.




(a) $\psi$ + CP + CW + SW + GP

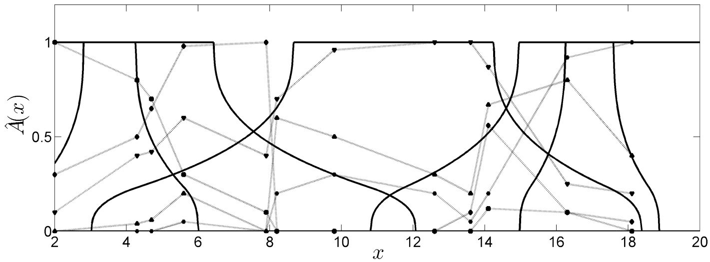



(b) CP + CW + SW + GP

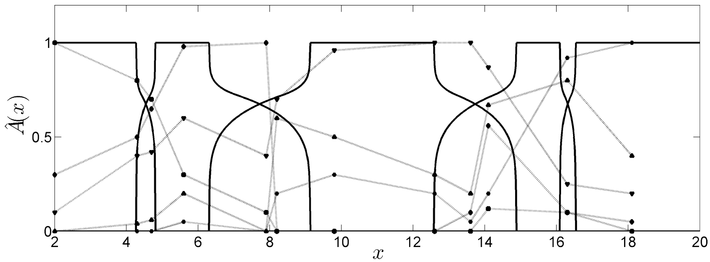



(c) $\psi$ + CW + SW + GP

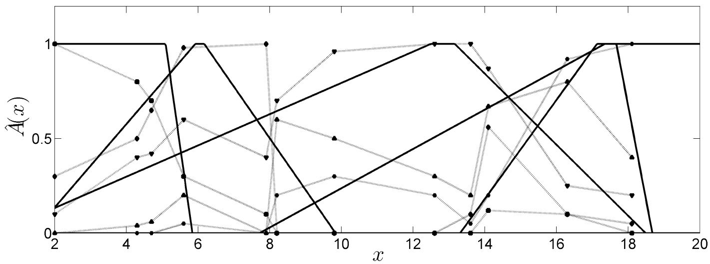
**Figure 41:** Typical partitions obtained by by SOGA $\psi$ + FM for the context-adapted fuzzy partition data set




(a) $\psi$ + CP + SW + GP




(b) $\psi$ + CP + CW + GP




(c) $\psi$ + CP + CW + SW

**Figure 42:** Typical partitions obtained by by SOGA $\psi$ + FM for the context-adapted fuzzy partition data set

**Table 7:** Results of the assessment of CA operators on the context-adapted fuzzy partition

| | | Operators | | | |
|---|---|---|---|---|---|
| $\psi$ | **CP** | **CW** | **SW** | **GP** | $MSE$ |
| ✔ | ✔ | ✔ | ✔ | ✔ | $0.0833 \pm 0.0059$ |
| ✘ | ✔ | ✔ | ✔ | ✔ | $0.1176 \pm 0.0032$ |
| ✔ | ✘ | ✔ | ✔ | ✔ | $0.1008 \pm 0.0003$ |
| ✔ | ✔ | ✘ | ✔ | ✔ | $0.0874 \pm 0.0126$ |
| ✔ | ✔ | ✔ | ✘ | ✔ | $0.1483 \pm 0.0045$ |
| ✔ | ✔ | ✔ | ✔ | ✘ | $0.0873 \pm 0.0073$ |
| ✔ | ✘ | ✘ | ✘ | ✘ | $0.1643 \pm 0.0077$ |
| ✘ | ✔ | ✘ | ✘ | ✘ | $0.2295 \pm 0.0000$ |
| ✘ | ✘ | ✔ | ✘ | ✘ | $0.2288 \pm 0.0000$ |
| ✘ | ✘ | ✘ | ✔ | ✘ | $0.1929 \pm 0.0000$ |
| ✘ | ✘ | ✘ | ✘ | ✔ | $0.1583 \pm 0.0000$ |
| ✘ | ✘ | ✘ | ✘ | ✘ | $0.2295 \pm 0.0000$ |

Table 7 shows the results of the assessment. In the Table, the operators used in each combination are identified by a tick. We repeated the SOGA five times for each combination and reported the mean $MSE \pm$ standard deviation computed as in Equation 6.2.

As expected, the combination which achieves the lowest $MSE$ is the one which comprises all the operators. Other combinations, which exclude one of more operators, perform worse even on this extremely simple data set. Indeed, the fuzzy modifiers were designed to be complementary with each other, and, therefore, their combined application allows to fully exploit the provided flexibility. Also, results highlight that, in this application, the most relevant contributions to adaptation are given by the non linear scaling function $\psi$ and by the support-width modifier. Remarkably, these results have been achieved in spite of the fact that the complete combination of operators is the one which determines the widest search space for the genetic learning process.

Figures 41 – 42 show typical context-adapted partitions obtained by some of the combinations evaluated in Table 7. Again, we can observe how the combined use of all of the operators (Figure 41(a)) provides an enhanced flexibility with respect to the other combinations (Figures 41(b) – 42(c)).

### 6.2.2 Single-Objective Genetic Algorithm

In this Section, we show and discuss the results of the application of SOGA $\psi$ + FM to the data sets described in Section 6.1.

#### 6.2.2.1 The Structure of Wages

We applied SOGA $\psi$ + FM to the structure of wages data set. For each of the four contexts and for each of the five folds, we executed five runs of the GA, with a population of $N_{\text{pop}} = 50$ individuals and a maximum of $T_{\max} = 100$ generations.

Since in Equation 5.9 we included a tunable parameter $\beta$ which controls the effects of the interpretability constraints on the fitness function, we repeated the runs of SOGA $\psi$ + FM with two different values of $\beta$, namely 0 and 0.1. We recall that setting $\beta = 0$ corresponds to performing an unconstrained search, i.e., to generating GFRBSs which do not necessarily enforce the interpretability constraints.

Table 8 shows the experimental results, expressed in terms of mean $MSE \pm$ standard deviation over the 25 runs performed for each context, both on the training and on the test set. As stated above, the same experiment was repeated on SOGA $\varphi_2$. The results of the simple AL $\varphi_0$ approach are also provided in the Table. Obviously, SOGA-based approaches always perform better than AL $\varphi_0$. We observe that SOGA $\psi$ + FM, $\beta = 0$ outperforms SOGA $\varphi_2$ in all contexts, except for the *college drop-outs* one. Actually, the SOGA $\varphi_2$ approach employs fewer bits in the chromosome (i.e., 50 against 154), and, therefore, it has a better chance to explore the search space for a global optimum. This is also testified by the standard deviations achieved by SOGA $\varphi_2$ on the train set, which are usually lower than the ones obtained by SOGA $\psi$ + FM. Further, we remark that,

**Table 8:** Results of the application of SOGA $\psi$ + FM to the structure of wages data set

| **Context** | **Approach** | $MSE_{train}$ | $MSE_{test}$ |
|---|---|---|---|
| *High school drop-outs* | SOGA $\psi$+FM, $\beta$ = 0 | 0.0340 $\pm$ 0.0033 | 0.0520 $\pm$ 0.0156 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0414 $\pm$ 0.0054 | 0.0568 $\pm$ 0.0071 |
| | SOGA $\varphi_2$ | 0.0413 $\pm$ 0.0054 | 0.0649 $\pm$ 0.0181 |
| | AL $\varphi_0$ | 0.3664 $\pm$ 0.0497 | 0.3581 $\pm$ 0.0894 |
| *High school graduates* | SOGA $\psi$+FM, $\beta$ = 0 | 0.0465 $\pm$ 0.0058 | 0.0632 $\pm$ 0.0146 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0509 $\pm$ 0.0070 | 0.0719 $\pm$ 0.0259 |
| | SOGA $\varphi_2$ | 0.0566 $\pm$ 0.0048 | 0.0794 $\pm$ 0.0299 |
| | AL $\varphi_0$ | 1.2227 $\pm$ 0.2539 | 1.1981 $\pm$ 0.2956 |
| *College drop-outs* | SOGA $\psi$+FM, $\beta$ = 0 | 0.0626 $\pm$ 0.0082 | 0.0877 $\pm$ 0.0233 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0685 $\pm$ 0.0130 | 0.0917 $\pm$ 0.0283 |
| | SOGA $\varphi_2$ | 0.0649 $\pm$ 0.0070 | 0.0736 $\pm$ 0.0295 |
| | AL $\varphi_0$ | 0.9852 $\pm$ 0.1208 | 0.9696 $\pm$ 0.2060 |
| *College graduates* | SOGA $\psi$+FM, $\beta$ = 0 | 0.0968 $\pm$ 0.0184 | 0.1339 $\pm$ 0.0118 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.1227 $\pm$ 0.0171 | 0.1579 $\pm$ 0.0269 |
| | SOGA $\varphi_2$ | 0.1247 $\pm$ 0.0086 | 0.1726 $\pm$ 0.0372 |
| | AL $\varphi_0$ | 1.7764 $\pm$ 0.1432 | 1.7519 $\pm$ 0.3055 |

though the results of SOGA $\psi$ + FM, $\beta = 0.1$ on the training sets are comparable to those of SOGA $\varphi_2$, the former approach is able to perform a better generalization than the latter one on the test sets.

#### 6.2.2.2 Parametric Function

The SOGA $\psi$ + FM was tested on the parametric function data set with $N_{\text{pop}} = 100$ and $T_{\max} = 200$. We repeated the experiment five times for each fold, as detailed in the previous Section. Further, to analyze how parameter $\beta$ affects the accuracy-interpretability trade-off, we run SOGA $\psi$ + FM with three different values of $\beta$, namely 0, 0.1, and 1.

Table 9 shows the aggregated results achieved by the five algorithms for each context, both on the training and on the test set. As already

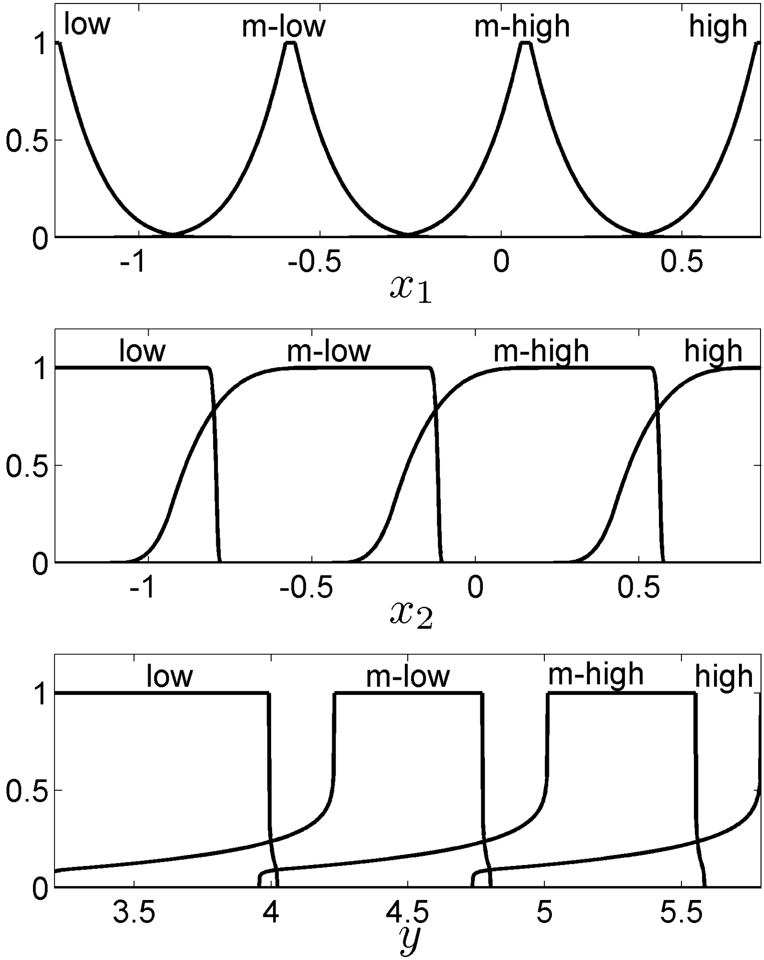



**Figure 43:** Typical partitions obtained by SOGA $\psi$ + FM, $\beta = 0$ for the parametric function data set in the $\kappa = 7$ context

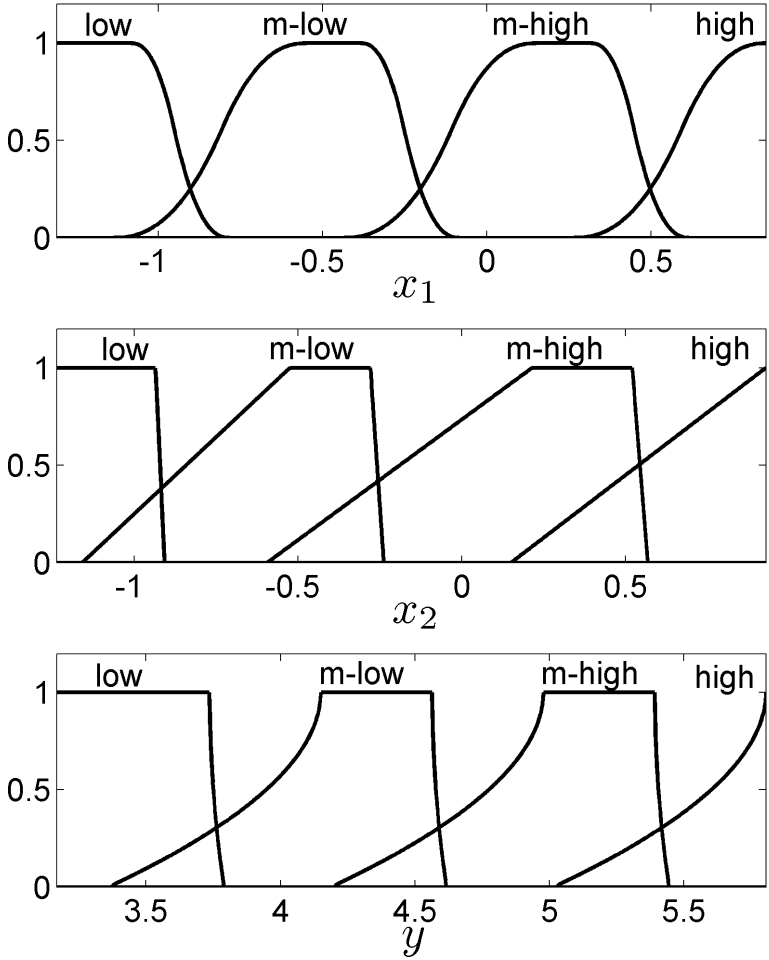



**Figure 44:** Typical partitions obtained by SOGA $\psi$ + FM, $\beta = 0.1$ for the parametric function data set in the $\kappa = 7$ context

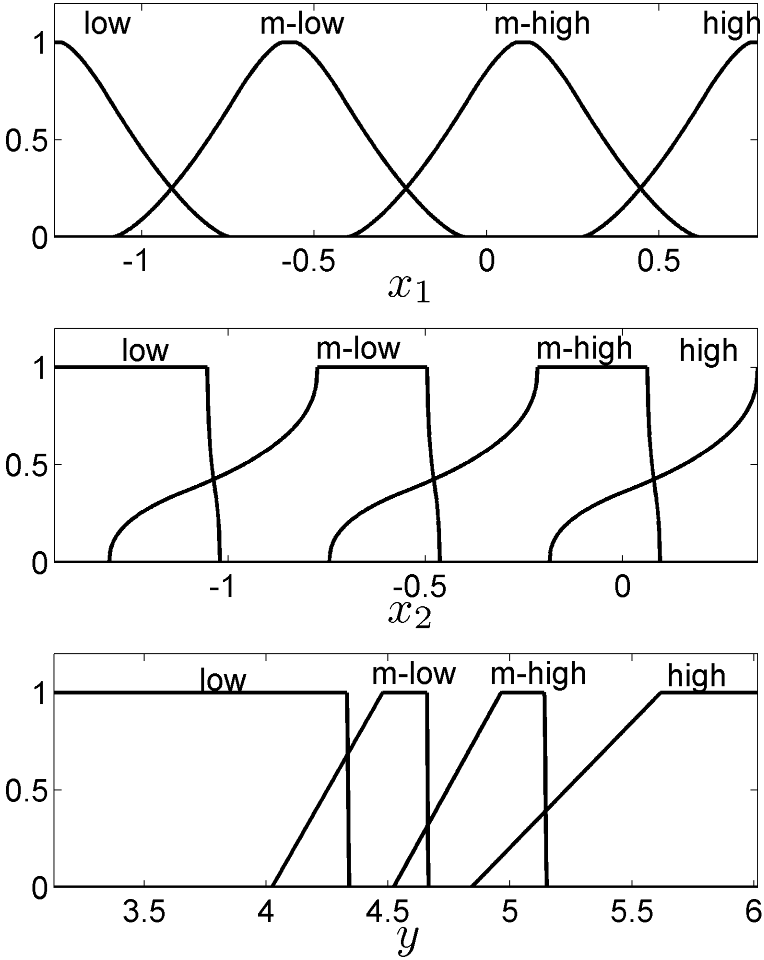



**Figure 45:** Typical partitions obtained by SOGA $\psi$ + FM, $\beta = 1$ for the parametric function data set in the $\kappa = 7$ context

**Table 9:** Results of the application of SOGA $\psi$ + FM to the parametric function data set

| Context | Approach | $MSE_{train}$ | $MSE_{test}$ |
|---|---|---|---|
| $\kappa = 2$ | SOGA $\psi$+FM, $\beta$ = 0 | 0.0048 $\pm$ 0.0007 | 0.0058 $\pm$ 0.0002 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0053 $\pm$ 0.0005 | 0.0069 $\pm$ 0.0004 |
| | SOGA $\psi$+FM, $\beta$ = 1 | 0.0059 $\pm$ 0.0002 | 0.0069 $\pm$ 0.0011 |
| | SOGA $\varphi_2$ | 0.0061 $\pm$ 0.0004 | 0.0070 $\pm$ 0.0013 |
| | AL $\varphi_0$ | 0.0341 $\pm$ 0.0012 | 0.0344 $\pm$ 0.0045 |
| $\kappa = 5$ | SOGA $\psi$+FM, $\beta$ = 0 | 0.0432 $\pm$ 0.0044 | 0.0566 $\pm$ 0.0150 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0477 $\pm$ 0.0035 | 0.0567 $\pm$ 0.0174 |
| | SOGA $\psi$+FM, $\beta$ = 1 | 0.0447 $\pm$ 0.0043 | 0.0580 $\pm$ 0.0180 |
| | SOGA $\varphi_2$ | 0.0442 $\pm$ 0.0031 | 0.0570 $\pm$ 0.0106 |
| | AL $\varphi_0$ | 0.0920 $\pm$ 0.0073 | 0.0919 $\pm$ 0.0295 |
| $\kappa = 7$ | SOGA $\psi$+FM, $\beta$ = 0 | 0.0666 $\pm$ 0.0060 | 0.0852 $\pm$ 0.0243 |
| | SOGA $\psi$+FM, $\beta$ = 0.1 | 0.0702 $\pm$ 0.0061 | 0.0895 $\pm$ 0.0275 |
| | SOGA $\psi$+FM, $\beta$ = 1 | 0.0710 $\pm$ 0.0065 | 0.0895 $\pm$ 0.0301 |
| | SOGA $\varphi_2$ | 0.0771 $\pm$ 0.0056 | 0.0937 $\pm$ 0.0234 |
| | AL $\varphi_0$ | 0.1616 $\pm$ 0.0101 | 0.1615 $\pm$ 0.0408 |

verified in the previous Section, SOGA $\psi$ + FM outperforms AL $\varphi_0$ and, generally, achieves better results than SOGA $\varphi_2$. Remarkably, the performance gaps with respect to SOGA $\varphi_2$ are particularly clear on the context determined by $\kappa = 7$, which, as it can be verified in Figure 38(c), is the most difficult to reproduce due to its extremely uneven shape. This is result can be obtained thanks to the modeling capabilities of the fuzzy modifiers introduced in Chapter 3. Finally, we observe that SOGA $\psi$ + FM achieves similar results for $\beta = 0.1$ and $\beta = 1$. Hence, in this example, small values of $\beta$ allow enforcing interpretability.

Figures 43 – 45 show typical partitions obtained by our CA technique using $\beta = 0$, $\beta = 0.1$, and $\beta = 1$ on the context determined by $\kappa = 7$. We observe that the initial uniform partitions with trapezoidal MFs are modified into non uniformly distributed partitions, with different fuzzy set shapes. As expected, the partitions obtained with $\beta = 0$ may show

Table 10: Results of the application of SOGA $\psi$ + FM to the fuel efficiency data set

| Context | Approach | $MSE_{train}$ | $MSE_{test}$ |
|---|---|---|---|
| *City* | SOGA $\psi$+FM, $\beta=0$ | $3.8821 \pm 0.3198$ | $2.2519 \pm 0.5463$ |
| | SOGA $\psi$+FM, $\beta=0.1$ | $4.2163 \pm 0.4437$ | $2.6471 \pm 0.7005$ |
| | SOGA $\psi$+FM, $\beta=1$ | $4.9980 \pm 0.4690$ | $3.4109 \pm 0.7581$ |
| | SOGA $\varphi_2$ | $5.6841 \pm 0.3119$ | $4.4560 \pm 0.9636$ |
| | AL $\varphi_0$ | $145.5427 \pm 33.5002$ | $151.2866 \pm 35.3240$ |
| *Highway* | SOGA $\psi$+FM, $\beta=0$ | $4.4373 \pm 0.3491$ | $3.8125 \pm 0.6276$ |
| | SOGA $\psi$+FM, $\beta=0.1$ | $4.8899 \pm 0.6598$ | $3.8739 \pm 0.6662$ |
| | SOGA $\psi$+FM, $\beta=1$ | $6.9441 \pm 0.5320$ | $6.7319 \pm 0.9976$ |
| | SOGA $\varphi_2$ | $7.4895 \pm 0.4978$ | $7.7359 \pm 2.0332$ |
| | AL $\varphi_0$ | $88.9757 \pm 35.0451$ | $94.0178 \pm 40.3049$ |

a too high (low) level of coverage with respect to the ones produced by $\beta = 1$, which tries to maintain the membership values of the crossing points in the $[\epsilon_{\min}, \epsilon_{\max}]$ interval (see Section 4.2).

#### 6.2.2.3 Fuel Consumption

The last test of SOGA $\psi$ + FM was performed on the fuel consumption data set. As in the previous applications, we performed five runs for each traffic condition and for each fold. In this experiment, we set $N_{\text{pop}} = 50$ and $T_{\max} = 50$.

Table 10 summarizes the results of the experiments. In this application, it can be noted that our SOGA $\psi$ + FM always outperforms the other approaches. Further, we can observe that the $MSE$ increases with the increase of $\beta$. Since low values of $\beta$ do not enforce interpretability, we expect that low values of $MSE$ correspond to low interpretability.

Figures 46 – 48 show the partitions obtained in a trial by our context adaptation technique for the city context with the three different values of $\beta$. In the Figures, we observe that the fuzzy sets which compose the partitions corresponding to $\beta = 1$ are highly distinguishable and, there-

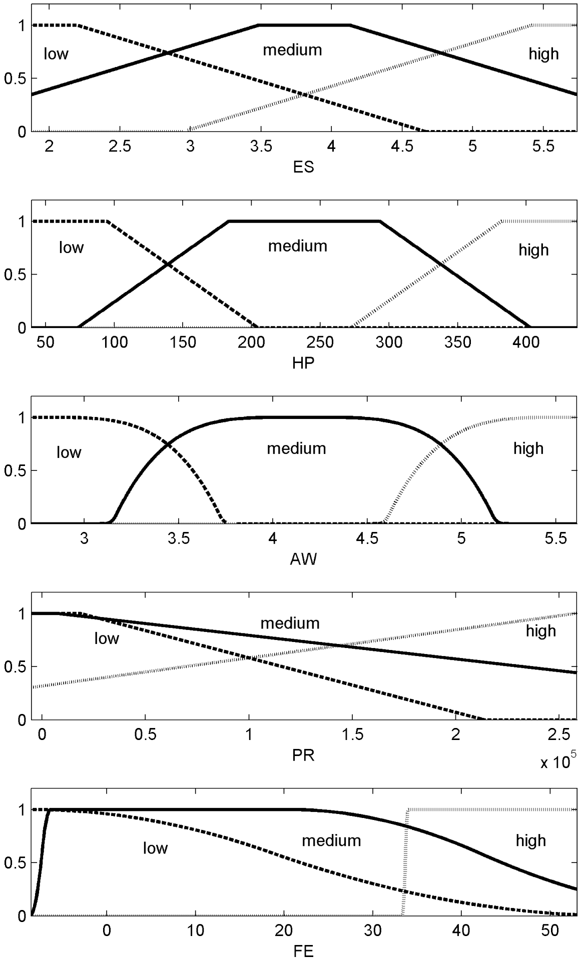



**Figure 46:** Typical partitions obtained by SOGA $\psi$ + FM for the fuel efficiency data set in the city context with $\beta = 0$

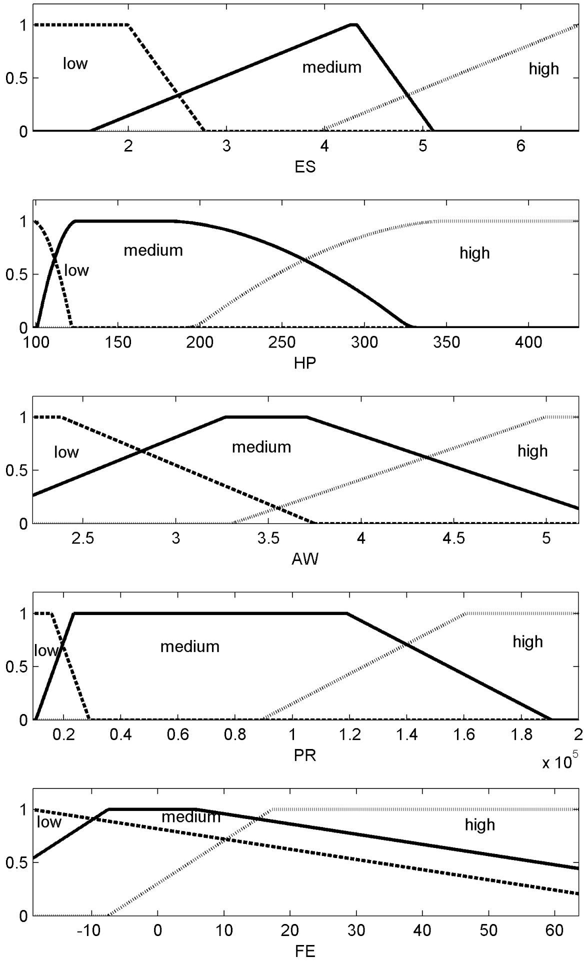



**Figure 47:** Typical partitions obtained by SOGA $\psi$ + FM for the fuel efficiency data set in the city context with $\beta = 0.1$

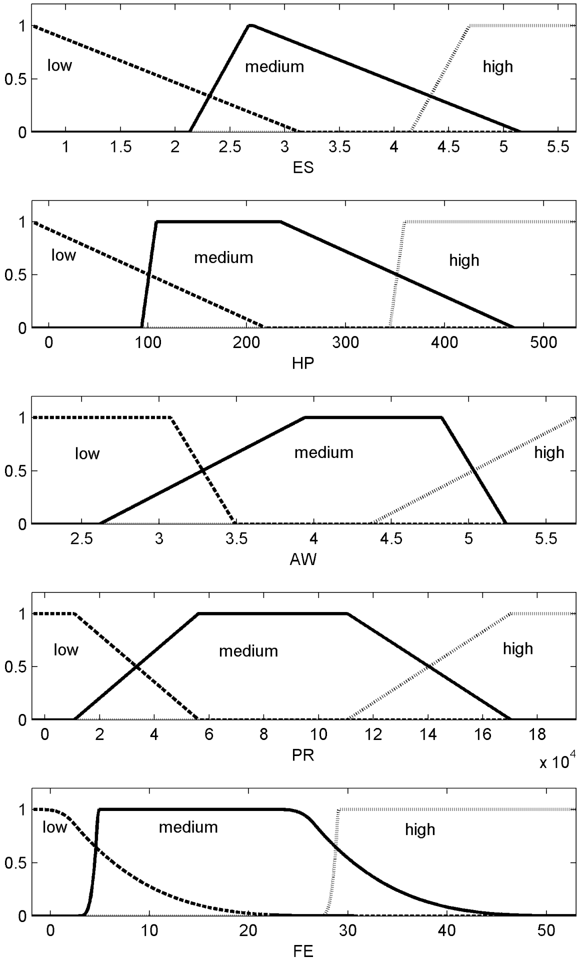



**Figure 48:** Typical partitions obtained by SOGA $\psi$ + FM for the fuel efficiency data set in the city context with $\beta = 1$

**Table 11:** Results of the application of MOEA $\psi$+FM to the structure of wages data set

| Context | Solution | $MSE_{train}$ | $MSE_{test}$ | $\bar{\Phi}_Y$ |
|---|---|---|---|---|
| *High school* | Best $MSE_{test}$ | 0.0500±0.0091 | 0.0531±0.0168 | 0.0815±0.0618 |
| *drop-outs* | Best $\bar{\Phi}_Y$ | 0.1919±0.1793 | 0.1895±0.1192 | 0.0059±0.0005 |
| *High school* | Best $MSE_{test}$ | 0.0523±0.0246 | 0.0488±0.0302 | 0.1649±0.0917 |
| *graduates* | Best $\bar{\Phi}_Y$ | 0.1919±0.1793 | 0.1895±0.1192 | 0.0062±0.0011 |
| *College* | Best $MSE_{test}$ | 0.0807±0.0550 | 0.0753±0.0440 | 0.1666±0.1242 |
| *drop-outs* | Best $\bar{\Phi}_Y$ | 0.8527±0.7593 | 0.7569±0.4536 | 0.0047±0.0015 |
| *College* | Best $MSE_{test}$ | 0.1167±0.0347 | 0.1312±0.0724 | 0.2266±0.1330 |
| *graduates* | Best $\bar{\Phi}_Y$ | 0.6888±0.5484 | 0.7061±0.7299 | 0.0048±0.0012 |

fore, guarantee a high degree of interpretability, despite a degradation of system accuracy. On the contrary, the fuzzy sets generated with $\beta = 0$ are less distinguishable and therefore difficultly interpretable. Finally, the fuzzy sets produced with the intermediate value $\beta = 0.1$ show a balance of the two properties. Obviously, the correct choice of $\beta$ is application-dependent: if the user is more interested in interpretability, he/she will choose high values of $\beta$; otherwise, if the user is more interested in performance, he/she will select values of $\beta$ close to 0.

### 6.2.3 Multi-Objective Evolutionary Algorithm

In this Section, we show and discuss the results of the application of MOEA $\psi$ + FM to the data sets described in Section 6.1.

#### 6.2.3.1 The Structure of Wages

We applied the MOEA $\psi$ + FM CA approach to the structure of wages data set with the same parameters employed in the experiments of Section 6.2.2.1, i.e., $N_{\text{pop}} = 50$ and $T_{\max} = 100$. To guarantee statistically valid results, the execution of NSGA-II was repeated for each of the five folds.

Table 11 shows the results of the application of MOEA $\psi$ + FM to the structure of wages data set. For each context, we reported the mean ±

standard deviation of $MSE_{train}$, $MSE_{test}$, and $\bar{\Phi}_Y$ for two selected GFRBSs, namely the ones which obtained the best $MSE_{test}$ and the best $\bar{\Phi}_Y$, respectively. These GFRBSs correspond to, respectively, the most accurate and the most interpretable among those comprised in the final Pareto front. We observe that the results obtained on the training set are comparable to those reported in Table 8. Further, the GFRBSs which obtain the best $MSE$s perform better than SOGA-based approaches on the test set. Hence, the MOEA $\psi$ + FM is less prone to overfitting than SOGA $\psi$ + FM and SOGA $\varphi_2$. Conversely, as expected, the GFRBSs which exhibit the best interpretability degrees are affected by a significant degradation of performances.

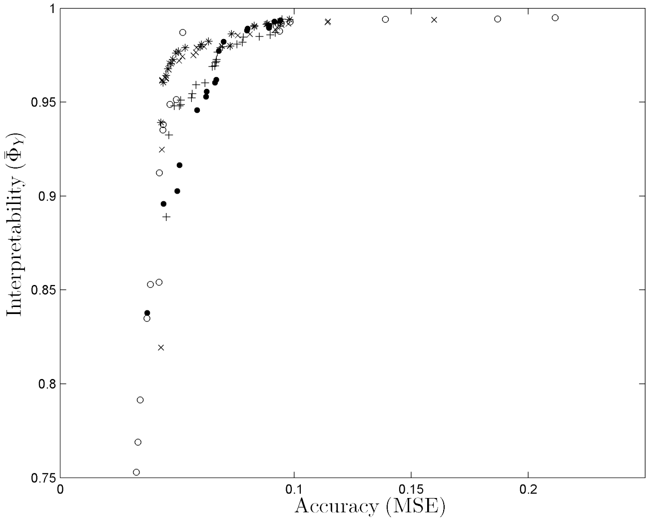
Figures 49 – 52 show the Pareto fronts obtained for each context on the train and test set. In the Figures, fronts obtained on the five folds are marked by different symbols. It can be noted that all the fronts on the training set are compact and well distributed. Also, we observe that the fronts obtained on the test set maintain the original shape determined by the accuracy-interpretability trade-off on the train set.

#### 6.2.3.2 Parametric Function

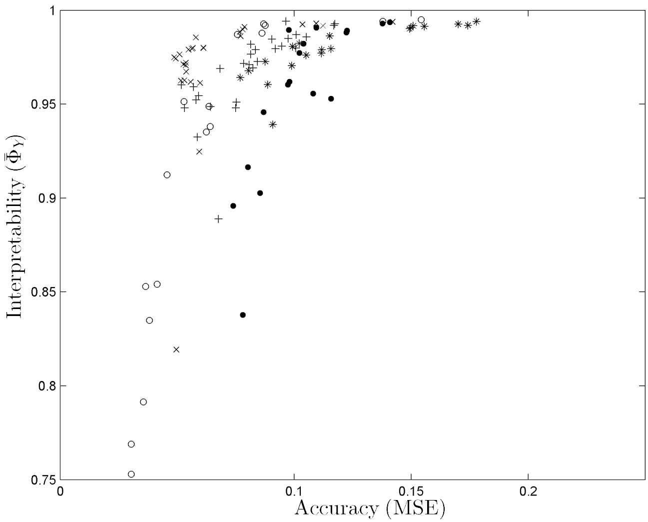
The application of MOEA $\psi$ + FM on the parametric function data set was performed with $N_{\text{pop}} = 50$ and $T_{\max} = 200$. Again, for each fold, we executed a run of the NSGA-II.

The MOEA $\psi$ + FM approach was compared with SOGA $\psi$ + FM, $\beta = 0$, and with SOGA $\varphi_2$. In both SOGAs, we adopted the same crossover and mutation probabilities as in NSGA-II. To ensure a fair comparison, the two SOGAs were tested with $N_{\text{pop}} = 50$ and $T_{\max} = 200$ as well. Moreover, to obtain statistically meaningful results, the experiments on SOGA $\psi$ + FM and SOGA $\varphi_2$ were repeated five times for each context and for each fold, as already performed in Section 6.2.2.

Table 12 shows the comparison among the results achieved by the three techniques taken into account. As regards MOEA $\psi$ + FM, we selected two GFRBSs: one, denoted as (a) in the Table, with the lowest $MSE_{test}$, and the other, denoted as (b), with the lowest $MSE_{test}$ among the solutions dominating the GFRBSs determined by SOGA $\varphi_2$. As ex-

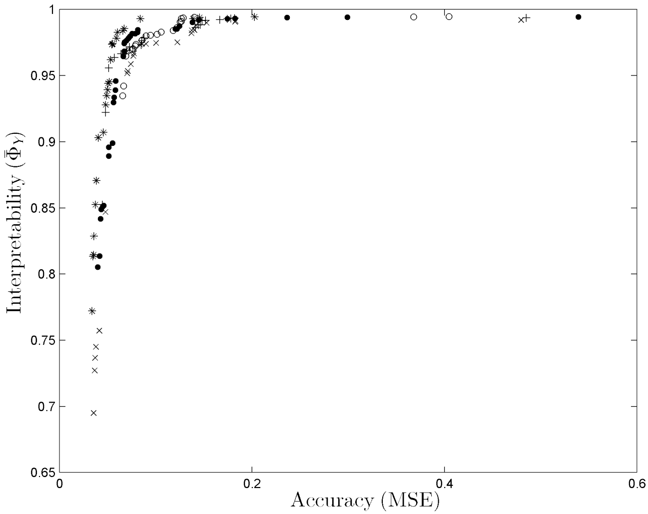



(a) Training set

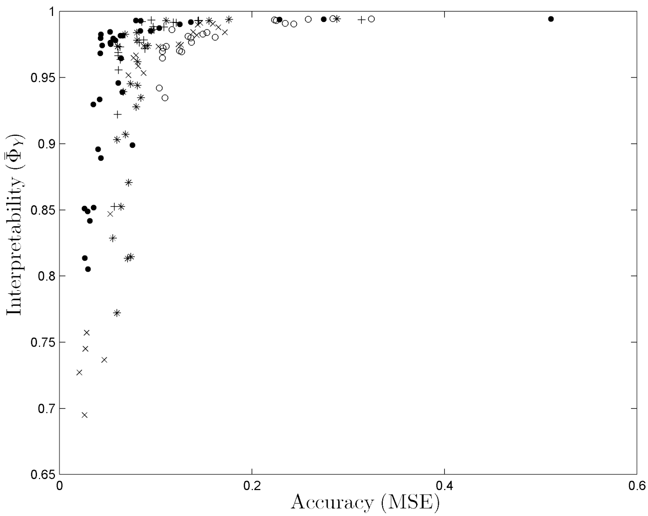



(b) Test set

**Figure 49:** Pareto fronts obtained by MOEA $\psi$ + FM for the structure of wages data set on the high school drop-outs context

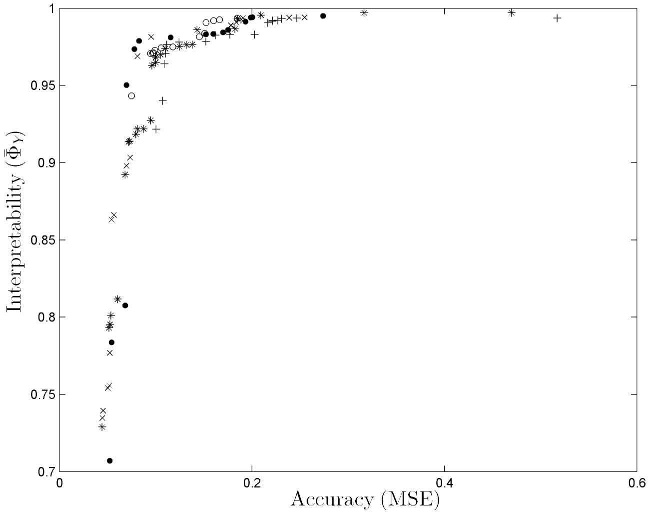



(a) Training set

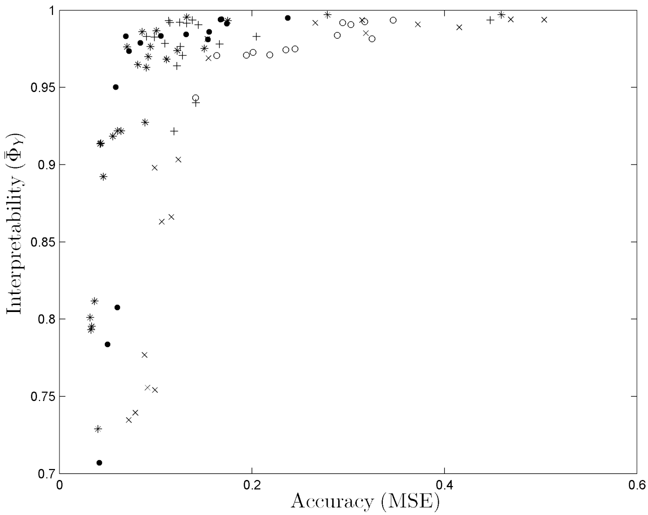



(b) Test set

**Figure 50:** Pareto fronts obtained by MOEA $\psi$ + FM for the structure of wages data set on the high school graduates context

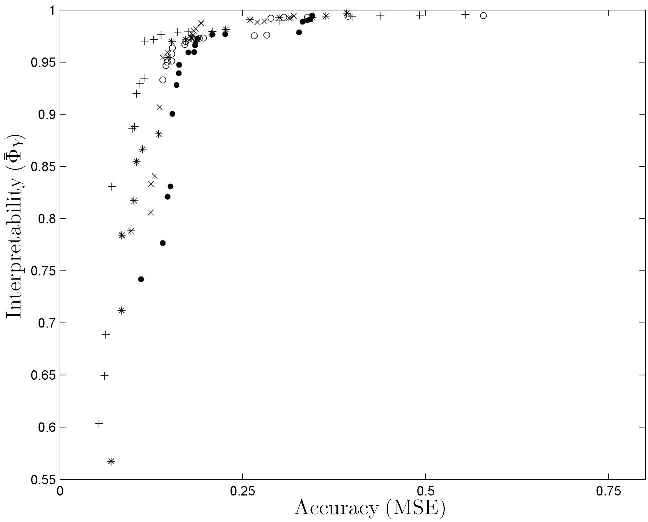



(a) Training set

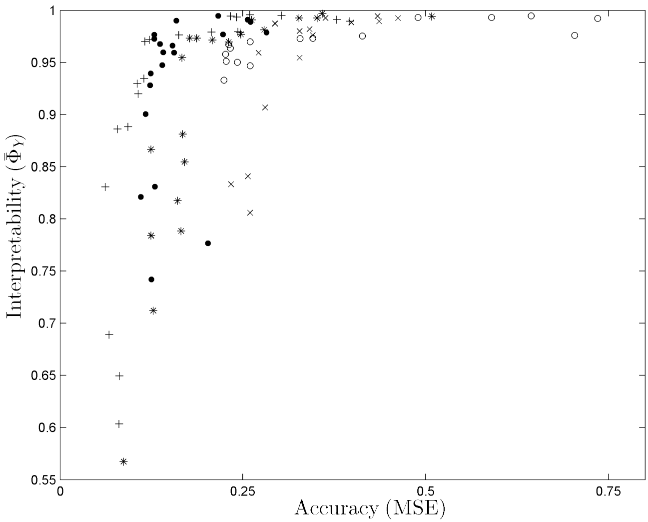



(b) Test set

**Figure 51:** Pareto fronts obtained by MOEA $\psi$ + FM for the structure of wages data set on the college drop-outs context




(a) Training set




(b) Test set

**Figure 52:** Pareto fronts obtained by MOEA $\psi$ + FM for the structure of wages data set on the college graduates context

**Table 12:** Results of the application of MOEA $\psi$+FM to the parametric function data set: (a-b) GFRBSs generated by MOEA $\psi$+FM, (c) SOGA $\psi$+FM, $\beta = 0$, (d) SOGA $\varphi_2$

| Context | Approach | $MSE_{train}$ | $MSE_{test}$ | $\bar{\Phi}_Y$ |
|---|---|---|---|---|
| $\kappa = 2$ | (a) | 0.0075 ± 0.0007 | 0.0066 ± 0.0017 | 0.9481 ± 0.0443 |
| | (b) | 0.0075 ± 0.0007 | 0.0066 ± 0.0017 | 0.9481 ± 0.0007 |
| | (c) | 0.0056 ± 0.0004 | 0.0068 ± 0.0016 | 0.7003 ± 0.0350 |
| | (d) | 0.0073 ± 0.0012 | 0.0094 ± 0.0032 | 0.8250 ± 0.0155 |
| $\kappa = 5$ | (a) | 0.0526 ± 0.0053 | 0.0509 ± 0.0139 | 0.9921 ± 0.0024 |
| | (b) | 0.0526 ± 0.0053 | 0.0509 ± 0.0139 | 0.9921 ± 0.0024 |
| | (c) | 0.0441 ± 0.0034 | 0.0554 ± 0.0111 | 0.7403 ± 0.0484 |
| | (d) | 0.0462 ± 0.0036 | 0.0575 ± 0.0153 | 0.8510 ± 0.0084 |
| $\kappa = 7$ | (a) | 0.0817 ± 0.0125 | 0.0811 ± 0.0333 | 0.9533 ± 0.0750 |
| | (b) | 0.0836 ± 0.0106 | 0.0812 ± 0.0334 | 0.9880 ± 0.0115 |
| | (c) | 0.0684 ± 0.0068 | 0.0842 ± 0.0273 | 0.6307 ± 0.0106 |
| | (d) | 0.0777 ± 0.0056 | 0.0918 ± 0.0250 | 0.8899 ± 0.0156 |

pected, the context-adapted GFRBSs generated by SOGA $\psi$ + FM on the test set are characterized by a low $MSE$, comparable, however, to the best $MSE$s obtained by the GFRBSs in the Pareto fronts. Nevertheless, their interpretability is poor, since the only objective of SOGA $\psi$ + FM with $\beta = 0$ is only to minimize the performance error. SOGA $\varphi_2$ achieves values of $MSE$ higher than SOGA $\psi$ + FM, but generates more interpretable GFRBSs, since it is a scaling function-based approach and, therefore, introduces a lower distortion in the fuzzy partitions than SOGA $\psi$ + FM. MOEA $\psi$ + FM provides the decision maker with a set of GFRBSs with different trade-offs between accuracy and interpretability. In particular, the solutions denoted as (a) and (b) achieve an $MSE$ equal to or lower than MOEA $\psi$ + FM and SOGA $\varphi_2$ on the test set, and are characterized by higher values of $\bar{\Phi}_Y$. Although SOGA $\varphi_2$ performs better than MOEA $\psi$ + FM on the training set, MOEA $\psi$ + FM outperforms it on test set. Indeed, due to its simplicity, SOGA $\varphi_2$ can perform a deeper exploration of the search space than MOEA $\psi$ + FM, but can also easily incur overfitting problems. On the other hand, MOEA $\psi$ + FM balances $MSE$ with interpretability and, therefore, is less prone to overfitting than SOGAs. Further, we observe that, as expected, solutions generated by SOGA $\psi$ + FM actually lie on a hypothetical extension of the Pareto front in a zone of low interpretability.

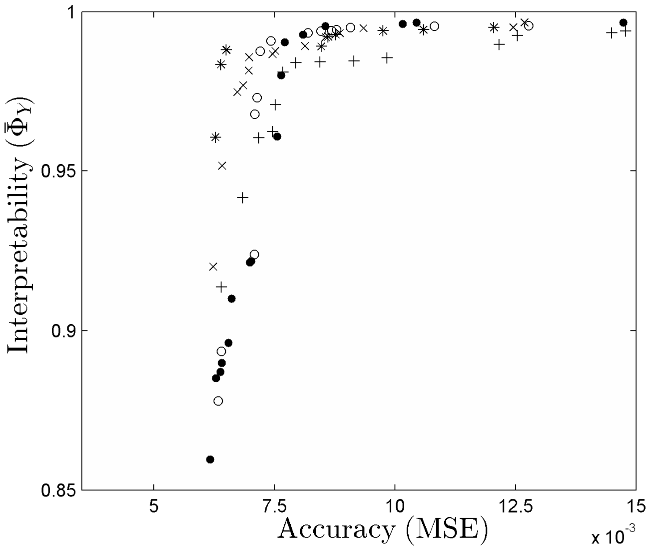
Figure 53 shows the Pareto fronts obtained by MOEA $\psi$ + FM on each of the five folds for the context $\kappa = 2$, both for the training and the test sets. We observe that the five fronts are quite wide and well distributed. Further, the fronts are all close to each other on the training set, thus highlighting that they do not depend on the particular execution of NSGA-II. Similar trends have been observed in the other contexts.

Figures 54 – 56 show, for the context corresponding to $\kappa = 2$, examples of the fuzzy partitions of the input and output variables of the GFRBSs generated by SOGA $\varphi_2$ (Figure 54) and SOGA $\psi$ + FM (Figure 55), and the fuzzy partitions of the GFRBS (a) which, in this context, corresponds also to (b) (Figure 56). We note that the GFRBS generated by SOGA $\psi$ + FM lacks coverage on $y$, whereas (a) shows a high interpretability degree and a low MSE.

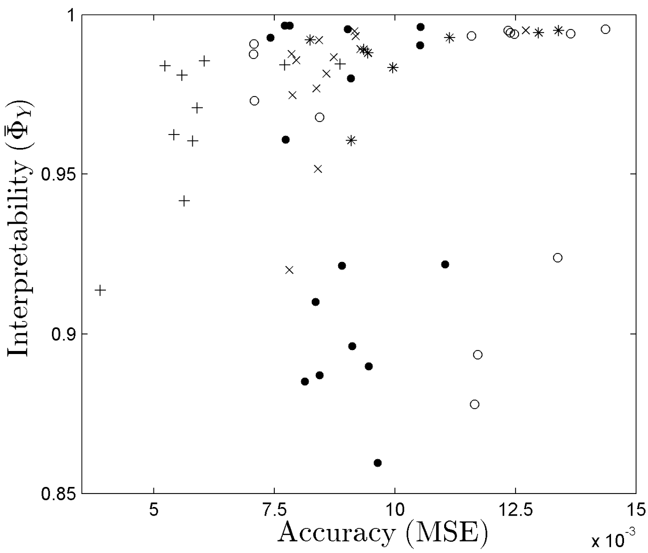
#### 6.2.3.3 Fuel Consumption

As in the previous example, we applied the MOEA $\psi$ + FM to the fuel consumption data set and compared it with the SOGA-based CA approaches. Again, we set $N_{\text{pop}} = 50$ and $T_{\max} = 200$ and we exploit the 5-fold cross-validation.

Table 13 summarizes the results obtained by comparing the three approaches. Figure 57 shows the Pareto fronts obtained for each fold and for both the training and test sets on the city context. As regards the highway context, similar fronts were obtained. Figure 57 and Table 13 confirm the trend that we observed in Section 6.2.3.2. Indeed, the GFRBSs determined by SOGA $\psi$ + FM are characterized by a low $MSE$ and a poor interpretability, while SOGA $\varphi_2$ generates GFRBSs with good trade-offs between accuracy and interpretability that are, however, Pareto-dominated by some solutions found by our MOEA. Further, on this data set, the GFRBSs identified by NSGA-II generalize much better than GFRBSs obtained




(a) Training set




(b) Test set

**Figure 53:** Pareto fronts obtained by MOEA $\psi$ + FM for the parametric function data set in the $\kappa = 2$ context

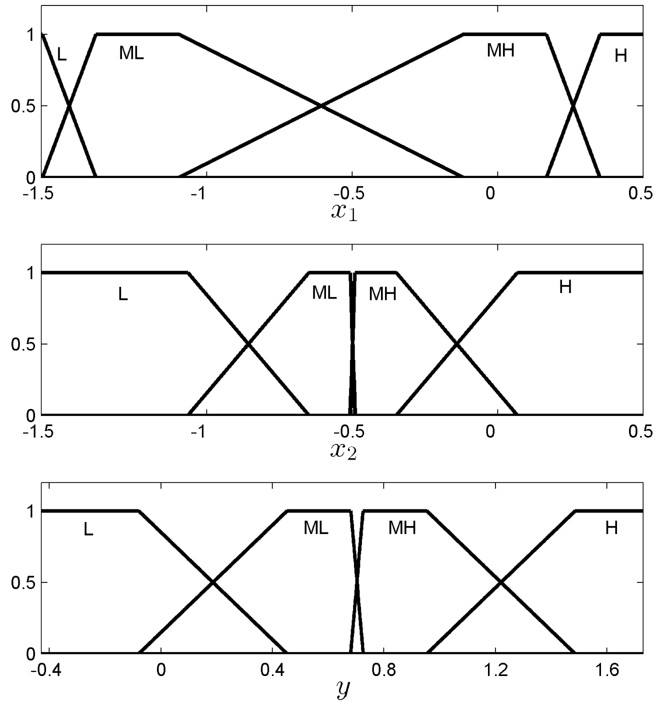



**Figure 54:** Typical partitions obtained by SOGA $\varphi_2$ for the parametric function data set in the $\kappa = 2$ context

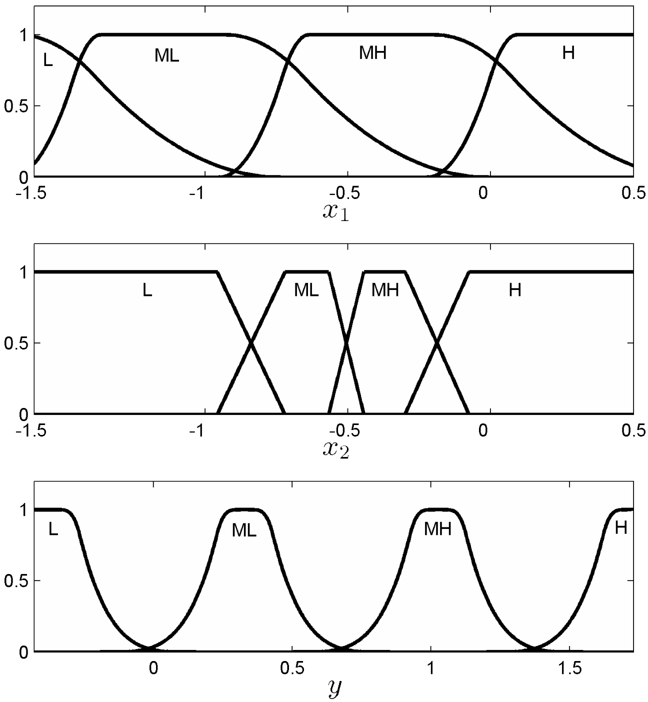



**Figure 55:** Typical partitions obtained by SOGA $\psi$ + FM, $\beta = 0$ for the parametric function data set in the $\kappa = 2$ context

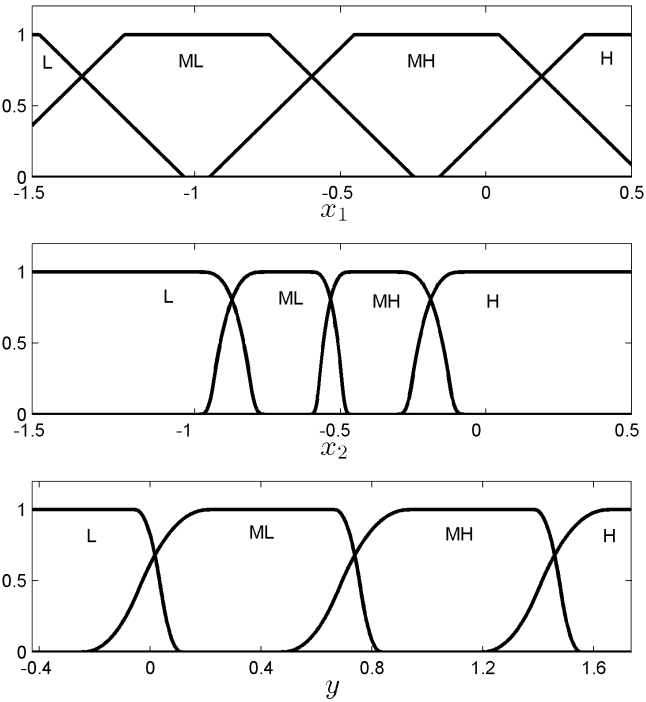


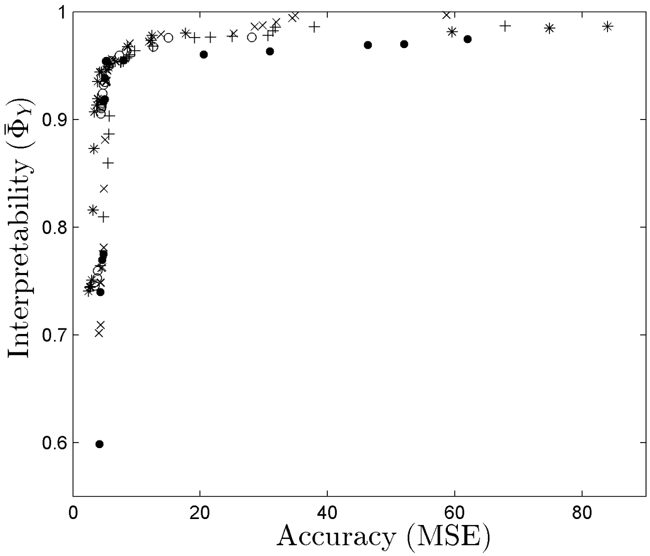

**Figure 56:** Typical partitions obtained by MOEA $\psi$ + FM for the parametric function data set in the $\kappa = 2$ context

**Table 13:** Results of the application of MOEA $\psi$+FM to the fuel efficiency data set: (a-b) GFRBSs generated by MOEA $\psi$+FM, (c) SOGA $\psi$+FM, $\beta = 0$, (d) SOGA $\varphi_2$

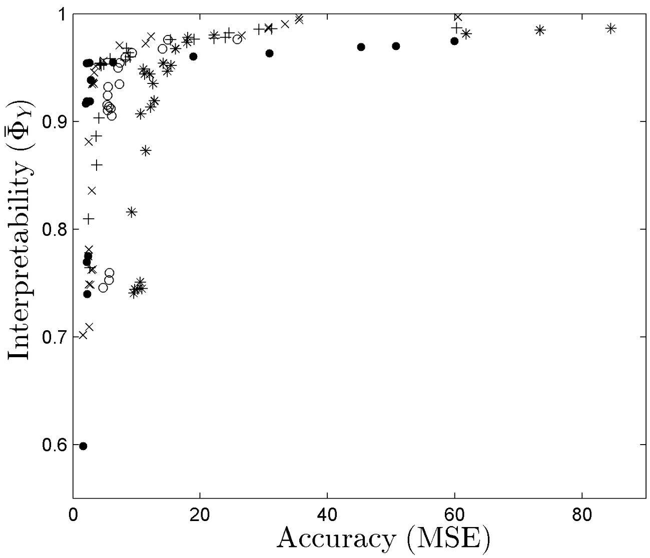
| Context | Approach | $MSE_{train}$ | $MSE_{test}$ | $\bar{\Phi}_Y$ |
|---|---|---|---|---|
| *City* | (a) | 3.6916 $\pm$ 0.7755 | 3.6821 $\pm$ 2.8287 | 0.7273 $\pm$ 0.0664 |
| | (b) | 4.6980 $\pm$ 0.8700 | 4.7619 $\pm$ 3.5105 | 0.8987 $\pm$ 0.0173 |
| | (c) | 3.0093 $\pm$ 0.5967 | 3.8554 $\pm$ 2.7852 | 0.4736 $\pm$ 0.0640 |
| | (d) | 5.0259 $\pm$ 0.4475 | 5.4802 $\pm$ 3.0011 | 0.8525 $\pm$ 0.0120 |
| *Highway* | (a) | 4.1016 $\pm$ 0.3193 | 4.6426 $\pm$ 1.9782 | 0.7740 $\pm$ 0.0402 |
| | (b) | 6.6064 $\pm$ 1.0363 | 6.3652 $\pm$ 2.4641 | 0.9229 $\pm$ 0.0365 |
| | (c) | 3.8348 $\pm$ 0.3534 | 4.6913 $\pm$ 1.7287 | 0.4778 $\pm$ 0.0940 |
| | (d) | 6.8556 $\pm$ 0.9315 | 7.6167 $\pm$ 2.0212 | 0.8731 $\pm$ 0.0108 |

by SOGA $\varphi_2$ and SOGA $\psi$ + FM, since they achieve similar $MSE$s on the training and test sets. This behavior can be explained by the set of CA operators adopted in our approach, which guarantees a higher modeling capability than the scaling function used in SOGA $\varphi_2$.

Figures 58 – 61 show, for the city context, sample fuzzy partitions of the input and output variables of the GFRBSs chosen as in the previous Section. We note that the partitions of the GFRBS generated by SOGA $\psi$ + FM (Figure 59), which outperforms the other in terms of accuracy on the training set, have interpretability difficulties. In particular, distinguishability among different fuzzy sets is not evident on the ES and the PR inputs. In contrast, the GFRBS (a) (Figure 60) achieves the lowest $MSE$ on the test set and it maintains its interpretability, even though the distinguishability of fuzzy sets on the PR input is not completely evident. Finally, the GFRBS (b) (Figure 61) outperforms all the other ones in terms of interpretability and, thanks to the modeling power of fuzzy modifiers, achieves an $MSE$ lower than SOGA $\varphi_2$.




(a) Training set




(b) Test set

**Figure 57:** Pareto fronts obtained for the fuel efficiency data set in the city context

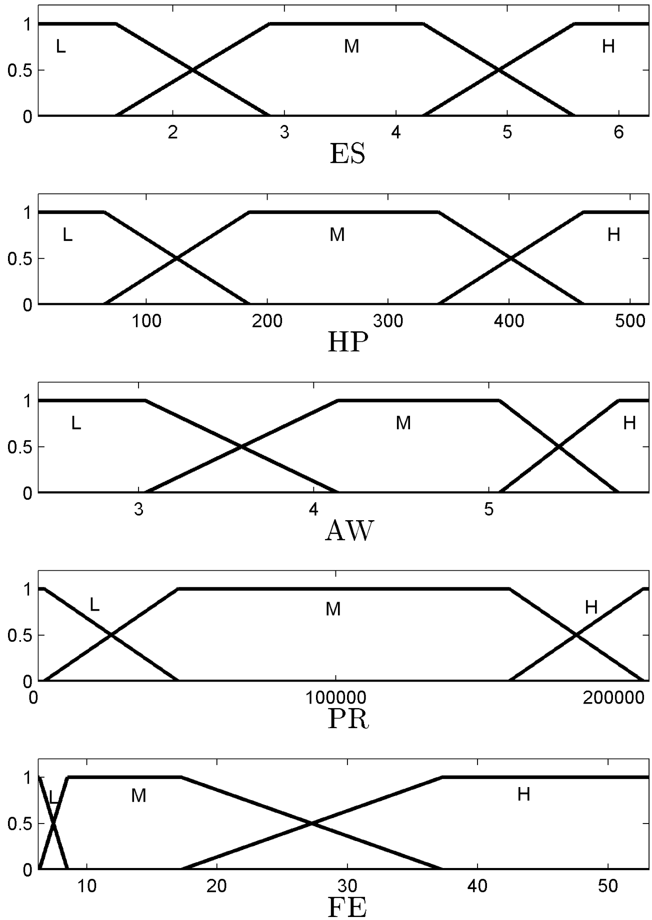



**Figure 58:** Typical partitions obtained by SOGA $\varphi_2$ for the fuel efficiency data set on the city context

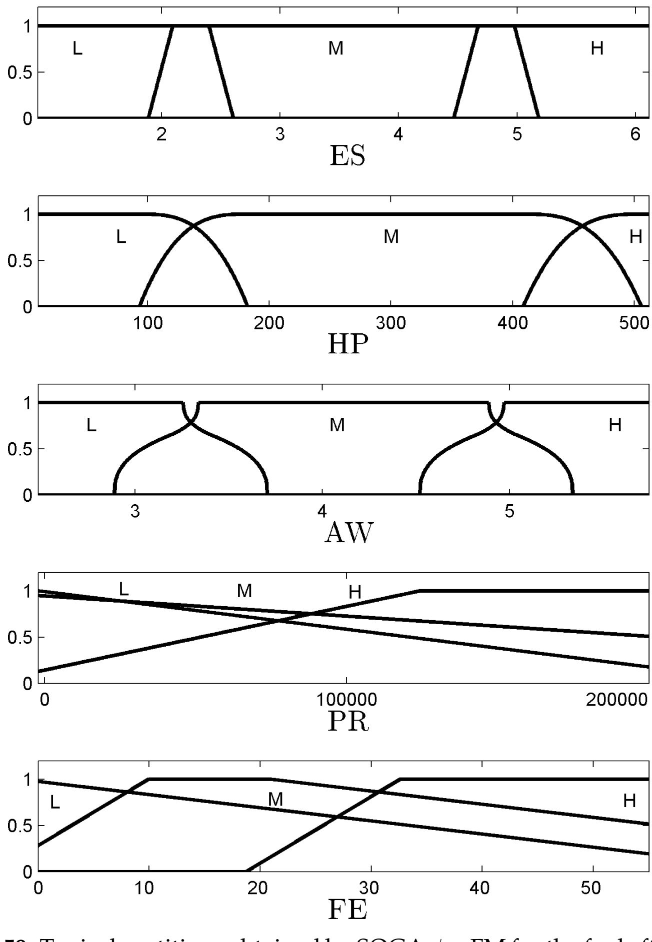



**Figure 59:** Typical partitions obtained by SOGA $\psi$ + FM for the fuel efficiency data set on the city context

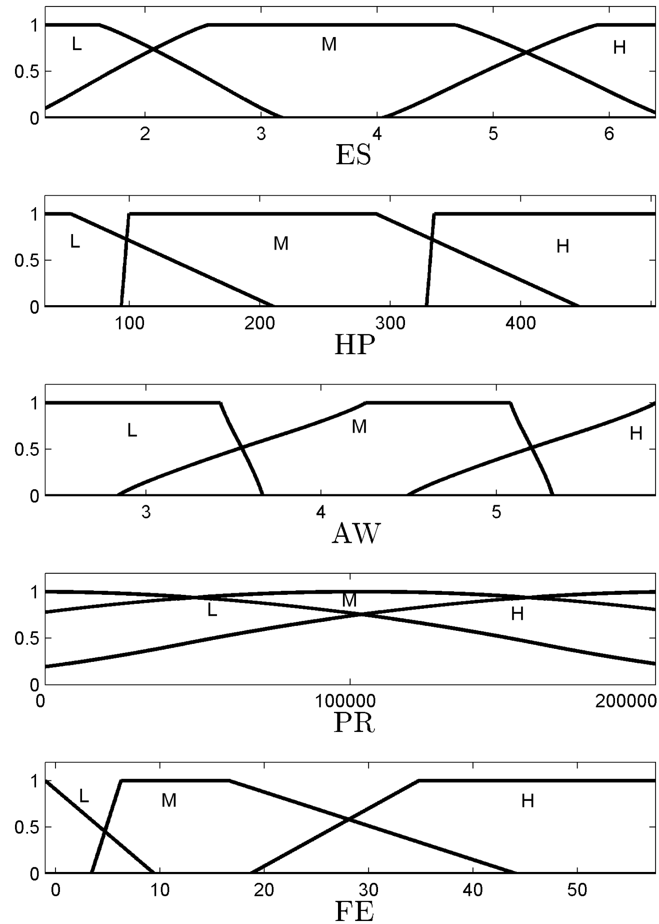



**Figure 60:** Typical partitions obtained by MOEA $\psi$ + FM (solution (a)) for the fuel efficiency data set on the city context

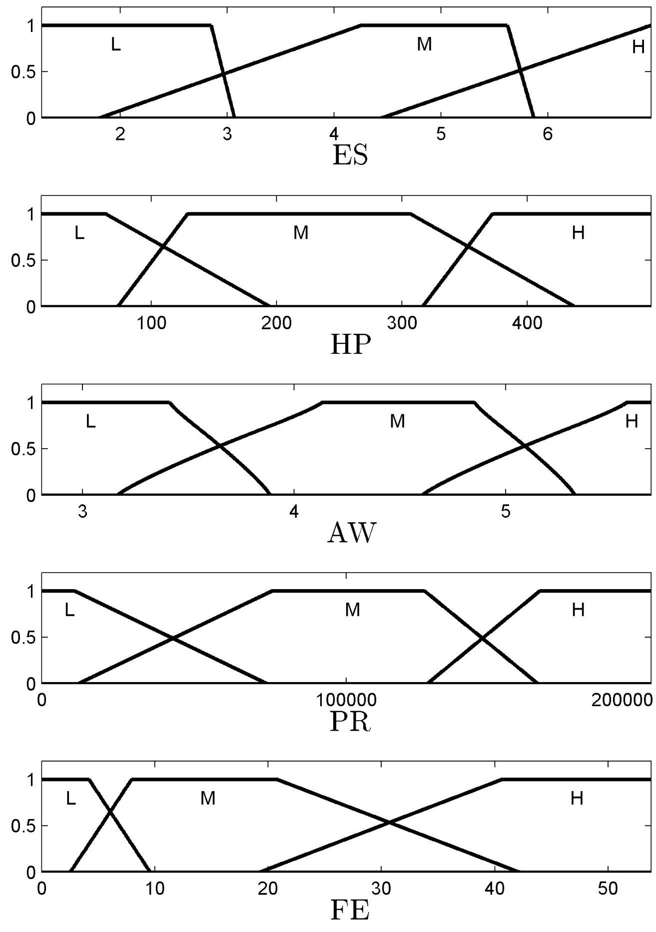



**Figure 61:** Typical partitions obtained by MOEA $\psi$ + FM (solution (b)) for the fuel efficiency data set on the city context

# Chapter 7

# Conclusion and Future Work

## 7.1 Conclusion

In this PhD thesis, we have deeply addressed the topic of CA of fuzzy systems. We started from the analysis and classification of previous approaches, which are mainly based on the use of scaling functions acting on the overall universe of discourse of normalized linguistic variables. Then, we introduced a conceptual framework of CA. We specialized our general proposal for CA of fuzzy systems, and, more precisely, of Mamdani-type FRBSs.

The RB of a fuzzy system expresses relations among linguistic terms which can be regarded as universal to a large extent and, therefore, independent of the specific application domain where the FRBS works. Conversely, the meanings associated with the linguistic terms generally depend upon the context and have to be adapted before being operative. Thus, once an FRBS has been identified on a given application domain, by knowledge elicitation or by abstraction, the same system can be used in a different domain, provided it is context-adapted by undergoing a proper instantiation or tuning process.

In line with existing techniques, we concentrated our research on the process of instantiation of a universal model to a context. To this aim, we introduced novel operators which allow an extremely flexible adaptation of fuzzy partitions. However, when augmenting the modeling capabilities of CA by using more sophisticated operators than simple scaling functions, techniques to deal with the preservation of interpretability must be taken into account. This issue has been addressed in the framework of the accuracy-interpretability trade-off of FRBSs. Hence, we proposed two interpretability indices properly suited for CA techniques, based, respectively, on evaluation of crossing points and on fuzzy ordering relations.

We remark that both the operators and the interpretability indices can be used in a wide range of automatic identification algorithms, not necessarily related to CA, as done, for instance, in (BDLM08), where the modifiers and $\Phi_Y$ are exploited in a cooperative coevolutionary algorithm for the identification of Mamdani-type FRBSs from data.

To perform the automatic instantiation of the universal model from contextualized data, we implemented two GFRBSs based on, respectively, a constrained GA and the NSGA-II. We extensively applied our algorithms to four data sets and compared them with other existing approaches. Results highlighted that techniques based on the combined application of our five operators usually perform better than simpler CA approaches, despite the larger search space that has to be explored by ours. The gap becomes particularly significant in case of slightly more complex data sets (like the fuel efficiency one) or when the context to be reproduced is extremely difficult (like $\kappa = 7$ in the parametric function data set). Once again, we remark that these results can be achieved thanks to the augmented modeling capabilities provided by our novel operators.

Finally, we summarize some of the most relevant theoretical and practical contributions developed in this work.

- A survey and a taxonomy of existing approaches to context adaptation of fuzzy systems.

- A conceptual reference framework for context adaptation.

- A set of guidelines for the development of instantiation algorithms for FRBSs.

- A novel flexible non linear scaling function in the line of previous approaches.

- A set of parametric orthogonal fuzzy modifiers.

- A novel interpretability index for fuzzy partitions based on fuzzy ordering relations.

- Two instantiation algorithms based on, respectively, a single- and a multi-objective GFRBS.

## 7.2 Future Work

Currently, the major drawback of the CA approach is that the RB is generated by knowledge elicitation, i.e., by including linguistic rules which are provided by human experts. In our conceptual framework, we described an alternative way to perform such task. Indeed, we believe that a promising research line related to CA of FRBSs consists in the study of efficient algorithms to perform abstraction of universal models from data collected in different contexts.

Such algorithms could be developed by leveraging previous work on information granules (Ped07), similarity (CS02), and RB learning by examples (WM92). Indeed, the abstraction task can be roughly divided in the following steps.

1. Identifying a vocabulary of similar linguistic terms (i.e., granules) shared by different contexts.

2. For each context, deriving an RB which comprises the terms of the vocabulary.

3. Merging all contextualized RBs in a context-free universal one by eliminating conflicting and insignificant rules.

Algorithms for the abstraction process would naturally fit in our conceptual framework for CA and, together with the different instantiation techniques described in this work, would provide a completely automatic approach to the generation of FRBSs from contextualized data.

Finally, a very interesting problem is the one related to modeling context in hierarchical architectures of FRBSs, such as the one proposed by Magdalena in (Mag02). Indeed, the majority of approaches to CA identifies a set of parameters that represent the effects of a given context on the linguistic variables. Obviously, for each new context, a different proper configuration of parameters have to be found. Unlike these approaches, hierarchical architectures try to capture the relations between the context variables and the set of parameters which characterize the effects of the context. In a sense, the context itself is modeled by an FRBS, whose outputs determine an on-line dynamic adaptation of the universal model.

To the best of our knowledge, currently in the literature there is no significative attempt for the development of learning algorithms for such hierarchical architectures. We believe that this kind of system could be of particular interest in the field of fuzzy control, i.e., in problems similar to the simple cart-pole balancing taken as example in Section 2.2. Indeed, in control systems, context variables can easily be identified and, further, on-line adaptation is typically required.

We performed some early studies to try to augment Magdalena's model with learning capability. Similarly to other hierarchical fuzzy systems, such as the one in (Wan99), we employed the backpropagation algorithm. However, our attempts did not show promising results because, due to the complex structure of the overall system, the learning signal cannot be properly backpropagated. On the other hand, approaches based on evolutionary algorithms do not seem suitable for such on-line learning because the search space grows quickly also for small problems.

Hence, we regard the automatic learning of hierarchical architectures of FRBSs for CA as a compelling problem with immediate practical applications.

# References

[Abr01] A. Abraham. Neuro fuzzy systems: State-of-the-art modeling techniques. In *Proceedings of the 6th International Work-Conference on Artificial and Natural Neural Networks (IWANN'01)*, pages 269–276, London, UK, 2001. 3, 16

[Bab02] R. Babuska. Fuzzy systems, modeling and identification, 2002. 14

[Bas94] A. Bastian. How to handle the flexibility of linguistic variables with applications. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 2(4):463–484, 1994. 28, 29, 30, 36, 40

[BB05] M. Bazire and P. Brézillon. Understanding context before using it. In A. K. Dey, B. N. Kokinov, D. B. Leake, and R. M. Turner, editors, *Proceedings of the 5th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2005)*, volume 3554 of *Lecture Notes in Computer Science*, pages 29–40, Paris, France, 2005. 22

[BBM07] D. Bacciu, A. Botta, and H. Melgratti. A fuzzy approach for negotiating quality of services. In *Proceedings of the 2nd Symposium on Trustworthy Global Computing (TCG 2006)*, volume 4661 of *Lecture Notes in Computer Sciences*, pages 200–217, Lucca, Italy, 2007. 13

[BDHA+07] P. Bosc, D. Dubois, A. Hadj-Ali, O. Pivert, and H. Prade. Adjusting the core and/or the support of a fuzzy set - a new approach to fuzzy modifiers. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'07)*, pages 1–6, London, UK, 2007. 49

[BDLM08] A. Botta, P. Ducange, B. Lazzerini, and F. Marcelloni. A multi-objective cooperative coevolutionary approach to Mamdani fuzzy system generation. In *Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2008)*, Malaga, Spain, 2008. 5, 21, 92, 139

[BKKP99] J. C. Bezdek, J. Keller, R. Krisnapuram, and N. R. Pal. *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer Academic, 1999. 3, 13

[BKP05] P. Bosc, D. Kraft, and F. Petry. Fuzzy sets in database and information systems: Status and opportunities. *Fuzzy Sets and Systems*, 156(3):418–426, 2005. 13

[BLM06a] A. Botta, B. Lazzerini, and F. Marcelloni. Context adaptation of Mamdani fuzzy systems through new operators tuned by a genetic algorithm. In *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'06)*, volume 1, pages 1641–1648, Vancouver, BC, Canada, 2006. 4, 5, 49, 86

[BLM06b] A. Botta, B. Lazzerini, and F. Marcelloni. New operators for context adaptation of Mamdani fuzzy systems. In D. Ruan, P. D'hondt, P. F. Fantoni, M. De Cock, M. Nachtegael, and E. E. Kerre, editors, *Applied Artificial Intelligence: Proceedings of the 7th Internations FLINS Conference*, pages 35–42. World Scientific, Genova, Italy, 2006. 4, 5, 49, 86

[BLM08] A. Botta, B. Lazzerini, and F. Marcelloni. Context adaptation of Mamdani fuzzy rule-based systems. *International Journal of Intelligent Systems*, 23(4):397 – 418, 2008. 4, 5, 49, 86

[BLMS07a] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Exploiting fuzzy ordering relations to preserve interpretability in context adaptation of fuzzy systems. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'07)*, pages 1–6, London, UK, 2007. 5, 92

[BLMS07b] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. A survey of approaches to context adaptation of fuzzy systems. In *Proceedings of the 10th IASTED International Conference on Intelligent Systems and Control (ISC 2007)*, Cambridge, MA, USA, 2007. 4

[BLMS08] A. Botta, B. Lazzerini, F. Marcelloni, and D. C. Stefanescu. Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index. *Soft Computing*, to appear, 2008. 4, 5, 49, 92

[Bor70] L. Borkowski, editor. *Selected works by J. Łukasiewicz*. North-Holland, 1970. 10

[Bré02] P. Brézillon. Modeling and using context: Past, present and future. Technical report, Laboratoire d'informatique de Paris 6, 2002. 22, 25

[Bru92] J. F. Brule. Fuzzy systems - A tutorial, 1992. 10

[BS01] J. G. Marìn Blàzquez and Q. Shen. Linguistic hedges on trapezoidal fuzzy sets: A revisit. In *Proceedings of the 2001 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'01)*, pages 412–415, 2001. 49

[BS03] J. G. Marìn Blàzquez and Q. Shen. Regaining comprehensebility of approximate fuzzy models via the use of linguistic hedges. In J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors, *Interpretability issues in fuzzy modeling*. Springer-Verlag, 2003. 16, 49, 55

[CCdJH05] J. Casillas, O. Cordón, M. J. del Jesus, and F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Transactions on Fuzzy Systems*, 13(1):13–29, 2005. 21

[CCHM03] J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors. *Interpretability issues in fuzzy modeling*. Springer-Verlag, 2003. 6, 16, 34, 67

[CdJH$^+$03] O. Cordón, M. J. del Jesus, F. Herrera, L. Magdalena, and P. Villar. A multiobjective genetic learning process for joint feature selection and granularity and contexts learning in fuzzy rule-based classification systems. In J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, editors, *Interpretability issues in fuzzy modeling*. Springer-Verlag, 2003. 28, 29, 32, 66, 67

[CDLM07] M. Cococcioni, P. Ducange, B. Lazzerini, and F. Marcelloni. A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. *Soft Computing*, 11(11):1013–1031, 2007. 16, 21

[CGH$^+$04] O. Cordón, F. A. C. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena. Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems*, 141:5–31, 2004. 3, 16, 20, 21, 31, 66, 84

[CHMV01] O. Cordón, F. Herrera, L. Magdalena, and P. Villar. A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. *Information Sciences*, 136(1–4), 2001. 4, 28, 29, 31, 32, 36, 40, 41, 106

[CK00] M. De Cock and E. E. Kerre. A new class of fuzzy modifiers. In *Proceedings of the 30th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2000)*, pages 121–126, Washington, DC, USA, 2000. 9, 49

[CK02] M. De Cock and E. E. Kerre. A context-based approach to linguistic hedges. *International Journal of Applied Mathematics and Computer Science*, 12(3):371–382, 2002. 6, 9, 35, 49

[CLM06] M. G. C. A. Cimino, B. Lazzerini, and F. Marcelloni. A novel approach to fuzzy clustering based on a dissimilarity relation extracted from data using a TS system. *Pattern Recognition*, 39(11):2077–2091, 2006. 12

[Coc01] M. De Cock. Fuzzy hedges: A next generation. In *Proceedings of the 13th European Summer School in Logic, Language and Information Student Session (ESSLLI'2001)*, pages 59–70, Helsinki, Finland, 2001. 6, 9, 49

[Coe02] C.A. Coello Coello. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. *Computer Methods in Applied Mechanics and Engineering*, 191(11–12):1245–1287, 2002. 85

[CS02] V. V. Cross and T. A. Sudkamp. *Similarity and compatibility in fuzzy set theory*. Springer-Verlag, 2002. 6, 67, 74, 140

[DAPM02] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transanctions on Evolutionary Computation*, 6(2):182–197, 2002. 19, 85, 92, 94

[DKLT05] A. K. Dey, B. N. Kokinov, D. B. Leake, and R. M. Turner, editors. *Proceedings of the 5th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2005)*, volume 3554 of *Lecture Notes in Computer Science*, Paris, France, 2005. 22

[DLJ00] D. Dumitrescu, B. Lazzerini, and L. C. Jain. *Fuzzy sets and their application to clustering and training*. CRC Press, 2000. 3, 6, 12

[dO99] J. V. de Oliveira. Semantic constraints for membership function optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 29(1):128–138, 1999. 6, 16, 56, 67, 73

[DW05] H. Deng and Y. Wang. An adaptive fuzzy logic controller with self-tuning scaling factors based on neural networks. In *Proceedings of the 7th International Symposium on Autonomous Decentralized Systems (ISADS 2005)*, pages 392–396, Chengdu, China, 2005. 31, 35

[DZG04] M. R. Delgado, F. Von Zuben, and F. A. C. Gomide. Coevolutionary genetic fuzzy systems: A hierarchical collaborative approach. *Fuzzy Sets and Systems*, 141:89–106, 2004. 21

[Flo05] R. V. Florian. Correct equations for the dynamic of the cart-pole system. Technical report, Center for Cognitive and Neural Studies (Coneural), 2005. 23

[GG94] R. R. Gudwin and F. A. C. Gomide. Context adaptation in fuzzy processing. In *Proceedings of the Brazil-Japan Joint Symposium on Fuzzy Systems*, Campinas, Brasil, 1994. 28, 29, 32, 34, 36, 40, 90, 106

[GGN98] R. R. Gudwin, F. A. C. Gomide, and M. A. Netto. A fuzzy elevator group controller with linear context adaptation. In *Proceedings of the 1998 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'98)*, pages 481–486, Anchorage, AK, USA, 1998. 34

[GGP98] R. Gudwin, F. A. C. Gomide, and W. Pedrycz. Context adaptation in fuzzy processing and genetic algorithms. *International Journal of Intelligent Systems*, 13(10–11):929–948, 1998. 26, 28, 32, 33, 34, 36, 41

[GRP+07] J. Gonzalez, I. Rojas, H. Pomares, L. J. Herrera, A. Guillen, J. M. Palomares, and F. Rojas. Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms. *International Journal of Approximate Reasoning*, 44(1):32–44, 2007. 16, 21

[Gui01] S. Guillaume. Designing fuzzy inference systems from data: An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9(3):426–443, 2001. 15, 16, 34, 67

[Hay94] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall, 1994. 3, 16

[Her08] F. Herrera. Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1(1):27–46, 2008. 3, 16, 20, 66, 84

[HHN02] V. N. Huynh, T. B. Ho, and Y. Nakamori. A parametric representation of linguistic hedges in Zadeh's fuzzy logic. *International Journal of Approximate Reasoning*, 30(3):203–223, 2002. 49

[HLV98] F. Herrera, M. Lozano, and J. L. Verdegay. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, 12(4):265–319, 1998. 17

[HN04] V. N. Huynh and Y. Nakamori. A context-based interpretation of linguistic variables. In *Proceedings of the 2004 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'04)*, volume 1, pages 257–262, Budapest, Hungary, 2004. 35

[IN07] H. Ishibuchi and Y. Nojima. Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, 44(1):4–31, 2007. 16, 21, 66, 85, 93

[Ish07] H. Ishibuchi. Multiobjective genetic fuzzy systems: Review and future research directions. In *Proceedings of the 2007 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'07)*, pages 1–6, London, UK, 2007. 16, 21, 84

[Jan93] J. S. R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–684, 1993. 3, 16

[JGSRB01] F. Jimenez, A. F. Gomez-Skarmeta, H. Roubos, and R. Babuska. Accurate, transparent, and compact fuzzy models for function approximation and dynamic modeling through multi-objective evolutionary optimization. In *Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization*, volume 2, pages 653–667, Zurich, Switzerland, 2001. 16

[Joh04] R. W. Johnson. New car and truck data, 2004. 101

[Jon06] K. A. De Jong. *Evolutionary computation: A unified approach*. MIT Press, 2006. 17

[Kar91] C. Karr. Genetic algorithms for fuzzy controllers. *AI Expert*, 6(2):26–33, 1991. 21, 23, 91

[KC00] J. D. Knowles and D. W. Corne. Approximating the nondominated front using the Pareto archived evolution strategy. *Evolutionary Computation*, 8(2):149–172, 2000. 19

[Kla06] F. Klawonn. Reducing the number of parameters of a fuzzy system using scaling functions. *Soft Computing*, 10(9):749–756, 2006. 28, 34, 36, 41

[KRRBV07] B. N. Kokinov, D. C. Richardson, T. Roth-Berghofer, and L. Vieu, editors. *Proceedings of the 7th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2007)*, volume 4635 of *Lecture Notes in Computer Science*, Roskilde, Denmark, 2007. 22

[KY95] G. J. Klir and B. Yuan. *Fuzzy sets and fuzzy logic - Theory and applications*. Prentice Hall, 1995. 3, 6, 10

[KY96] G. J. Klir and B. Yuan, editors. *Fuzzy sets, fuzzy logic, and fuzzy systems - Selected papers by L. A. Zadeh*. World Scientific, 1996. 3, 6, 14

[MA75] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975. 13, 14

[Mag97] L. Magdalena. Adapting the gain of an FLC with genetic algorithms. *International Journal of Approximate Reasoning*, 17(4):327–349, 1997. 28, 31, 33, 36, 40

[Mag02] L. Magdalena. On the role of context in hierarchical fuzzy controllers. *International Journal of Intelligent Systems*, 17(5):471–493, 2002. x, 26, 28, 35, 36, 91, 141

[Mam76] E. H. Mamdani. Advances in linguistic synthesis of fuzzy controllers. *International Journal of Man-Machine Studies*, 8:669–678, 1976. 6, 12

[Mat06] The MathWorks. Matlab R2006a, 2006. 64, 105

[MCFB04] C. Mencar, G. Castellano, A. M. Fanelli, and A. Bargiela. Similarity vs. possibility in measuring fuzzy sets distinguishability. In *Proceedings of 5th International Conference on Recent Advances in Soft Computing (RASC 2004)*, Nottingham, UK, 2004. 67, 68

[Men04] C. Mencar. *Theory of fuzzy information granulation: Contributions to interpretability issues*. PhD thesis, Department of Informatics, University of Bari, 2004. 67

[Mic99] Z. Michalewicz. *Genetic algorithms + data structures = evolution programs*. Springer-Verlag, 1999. 3, 17, 89, 91

[MJG05] R. Mikut, J. Jäkel, and L. Gröll. Interpretability issues in data-based learning of fuzzy systems. *Fuzzy Sets and Systems*, 150(2):179–197, 2005. 16

[MP99] R. K. Mudi and N. R. Pal. A robust self-tuning scheme for PI- and PD-type fuzzy controllers. *IEEE Transactions on Fuzzy Systems*, 7(1):2–16, 1999. 31

[MS96] Z. Michalewicz and M. Schoenauer. Evolutionary algorithms for constrained parameter optimization problems. *Evolutionary Computation*, 4(1):1–32, 1996. 17, 85

[MW92] K. M. Murphy and F. Welch. The structure of wages. *Quarterly Journal of Economics*, 107(1):285–326, 1992. 97

[Nau03] D. D. Nauck. Measuring interpretability in rule-based classification systems. In *Proceedings of the 2003 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'03)*, volume 1, pages 196–201, 2003. 16

[PD04] R. P. Paiva and A. Dourado. Interpretability and learning in neuro-fuzzy systems. *Fuzzy Sets and Systems*, 147(1):17–38, 2004. 16

[Ped07] W. Pedrycz. Granular computing - The emerging paradigm. *Journal of Uncertain Systems*, 1(1):38–61, 2007. 140

[PGG97] W. Pedrycz, R. R. Gudwin, and F. A. C. Gomide. Nonlinear context adaptation in the calibration of fuzzy sets. *Fuzzy Sets and Systems*, 88(1):91–97, 1997. 22, 25, 28, 32, 33, 34, 36, 40, 96

[PJ00] M. A. Potter and K. A. De Jong. Cooperative coevolution: An architecture for evolving coadapted subcomponents. *Evolutionary Computation*, 8(1):1–29, 2000. 19

[PRS01] C. A. Peña-Reyes and M. Sipper. Fuzzy CoCo: A cooperative-coevolutionary approach to fuzzy modeling. *IEEE Transactions on Fuzzy Systems*, 9(5):717–737, 2001. 21

[PV99] W. Pedrycz and A. V. Vasilakos. Linguistic models and linguistic modeling. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 29(6):745–757, 1999. 13

[Ros04] T. J. Ross. *Fuzzy logic with engineering applications*. Wiley, 2004. 3, 6, 10, 14, 30, 31, 34, 74

[RP01] M. Reformat and W. Pedrycz. Evolutionary optimization of information granules. In *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pages 2035–2040, Vancouver, BC, Canada, 2001. 21

[SBKvNL98] M. Setnes, R. Babuska, U. Kaymak, and H. R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 28(3):376–386, 1998. 67

[Sch04] R. Schulz. A Matlab implementation of NSGA-II, 2004. 105

[SR00] M. Setnes and H. Roubos. GA-fuzzy modeling and classification: complexity and performance. *IEEE Transactions on Fuzzy Systems*, 8(5):509–522, 2000. 21, 67

[SWK01] H. Shi, R. Ward, and N. Kharma. Expanding the definitions of linguistic hedges. In *Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference*, pages 2591–2595, Vancouver, BC, Canada, 2001. 49, 58

[Tur97] R. M. Turner. Determining the context-dependent meaning of fuzzy subsets. In *Proceedings of the 1997 International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-97)*, Rio de Janeiro, Brasil, 1997. 22, 34, 36

[Tur98] R. M. Turner. Context-mediated behavior for intelligent agents. *International Journal of Human-Computer Studies*, 48(3):307–330, 1998. 34, 36

[Wan97] X. Wang. An investigation into relations between some transitivity-related concepts. *Fuzzy Sets and Systems*, 89(2):257–262, 1997. 74, 76

[Wan99] L.-X. Wang. Analysis and design of hierarchical fuzzy systems. *IEEE Transactions on Fuzzy Systems*, 7(5):617 – 624, 1999. 141

[WCL00] Z.-W. Woo, H.-Y. Chung, and J.-J. Lin. A PID type fuzzy controller with self-tuning scaling factors. *Fuzzy Sets and Systems*, 115(2):321–326, 2000. 31

[WK01a] X. Wang and E. E. Kerre. Reasonable properties for the ordering of fuzzy quantities (i). *Fuzzy Sets and Systems*, 118(3):375–385, 2001. 74

[WK01b] X. Wang and E. E. Kerre. Reasonable properties for the ordering of fuzzy quantities (ii). *Fuzzy Sets and Systems*, 118(3):387–405, 2001. 74, 77, 81

[WKJ+05] H. Wang, S. Kwong, Y. Jin, W. Wei, and K.-F. Man. Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. *Fuzzy Sets and Systems*, 149(1):149–186, 2005. 16, 21

[WM92] L.-X. Wang and J. M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(6):1414–1427, 1992. 36, 69, 106, 140

[XI05] R. Xu and D. Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005. 12

[Yua91] Y. Yuan. Criteria for evaluating fuzzy ranking methods. *Fuzzy Sets and Systems*, 43(2):139–157, 1991. 77, 81

[Zad65] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965. 1, 6

[Zad73] L. A. Zadeh. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(1):28–44, 1973. 6, 58

[Zad94] L. A. Zadeh. Soft computing and fuzzy logic. *IEEE Software*, 11(6):48–56, 1994. 3, 6

[Zad96] L. A. Zadeh. Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4(2):103–111, 1996. 2

[ZYYLY$^+$06] X. Zong-Yi, H. Yuan-Long, Z. Yong, J. Li-Min, and H. Yuexian. A multi-objective cooperative coevolutionary algorithm for constructing accurate and interpretable fuzzy systems. In *Proceedings of the 2006 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'06)*, pages 1404–1410, Vancouver, BC, Canada, 2006. 21, 67