

# Context adaptation of fuzzy systems through a multi-objective evolutionary approach based on a novel interpretability index

Alessio Botta · Beatrice Lazzerini ·  
Francesco Marcelloni · Dan C. Stefanescu

Published online: 2 September 2008  
© Springer-Verlag 2008

**Abstract** Context adaptation (CA) based on evolutionary algorithms is certainly a promising approach to the development of fuzzy rule-based systems (FRBSs). In CA, a context-free model is instantiated to a context-adapted FRBS so as to increase accuracy. A typical requirement in CA is that the context-adapted system maintains the same interpretability as the context-free model, a challenging constraint given that accuracy and interpretability are often conflicting objectives. Furthermore, interpretability is difficult to quantify because of its very nature of being a qualitative concept. In this paper, we first introduce a novel index based on fuzzy ordering relations in order to provide a measure of interpretability. Then, we use the proposed index and the mean square error as goals of a multi-objective evolutionary algorithm aimed at generating a set of Pareto-optimum context-adapted Mamdani-type FRBSs with different trade-offs between accuracy and interpretability. CA is obtained through the use of specifically designed operators that adjust the universe of the input and

output variables, and modify the core, the support and the shape of fuzzy sets characterizing the partitions of these universes. Finally, we show results obtained by using our approach on synthetic and real data sets.

**Keywords** Fuzzy rule-based systems · Context adaptation · Multi-objective evolutionary algorithms · Fuzzy partition interpretability

## 1 Introduction

In the last years, Mamdani-type FRBSs (Mamdani and Assilian 1975) have been extensively and successfully applied to several engineering domains such as control, classification, regression and identification. As stated by Cordon et al. (2004), Mamdani-type FRBSs consist of:

- A *rule base* (RB) composed of linguistic IF-THEN rules, such as *if the temperature is hot then the fan speed is high*, where both the antecedent and the consequent parts are fuzzy propositions.
- A *data base* (DB), which associates a semantics, represented by means of fuzzy sets, with the linguistic terms used in the RB, e.g., *hot* and *high*.

The RB is often derived from heuristic knowledge, which is usually valid independently of the real environment where the FRBS will work. Thus, the RB can be considered as a context-free model. Indeed, the real environment does not affect the RB, but rather influences the DB, and, more specifically, the meaning associated with each linguistic term used in the rules. For instance, in the aforementioned rule, the meaning associated with the linguistic term *high* depends on the heat perception of each person, which is biased by, e.g., the latitude.

---

A. Botta (✉)  
IMT Lucca Institute for Advanced Studies,  
Piazza San Ponziano 6, 55100 Lucca, Italy  
e-mail: a.botta@imtlucca.it

B. Lazzerini · F. Marcelloni  
Dipartimento di Ingegneria dell'Informazione: Elettronica,  
Informatica, Telecomunicazioni, University of Pisa,  
via Diotisalvi 2, 56122 Pisa, Italy  
e-mail: b.lazzerini@iet.unipi.it

F. Marcelloni  
e-mail: f.marcelloni@iet.unipi.it

D. C. Stefanescu  
Mathematics and Computer Science Department,  
Suffolk University, Beacon Hill, Boston, MA, USA  
e-mail: dan@mcs.suffolk.edu

The several interesting possible applications have raised a large interest among researchers and practitioners in developing effective and efficient methods to generate highly performing and/or interpretable Mamdani-type FRBSs. In this framework, context adaptation (CA) is one of the most promising techniques, though it cannot be considered completely explored yet. CA of FRBSs moves from the following assumption: the RB models *universal knowledge* that can be reused in many different environments (Pedrycz et al. 1997). In each environment, the meanings associated with linguistic terms have to be adjusted so as to improve accuracy of the FRBS. Thus, CA is a tuning process that exploits context-specific information so as to adapt the parameters contained in the DB (Cordón et al. 2004).

The above-mentioned assumption implies that the RB should not be modified during the adaptation process. Since there exists a semantic relationship between the RB and the DB, this implication forces some constraints on the tuning 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 vary 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 the fuzzy sets should reflect the pre-CA ordering of linguistic terms.

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

1. Context adaptation should not modify the RB;
2. Context adaptation should not change the number of linguistic terms defined in the RB and, consequently, the number of corresponding fuzzy sets;
3. Context adaptation should not affect the semantic ordering of linguistic terms.

In this paper, we propose an approach to CA of Mamdani-type FRBSs aimed at improving accuracy while following the aforementioned guidelines and preserving interpretability of partitions (de Oliveira 1999; Casillas et al. 2003). Our approach exploits a novel interpretability index based on fuzzy ordering relations. The proposed index and the MSE are used as objectives of a MOEA aimed at generating a set of Pareto-optimum FRBSs with different trade-offs between accuracy and interpretability.

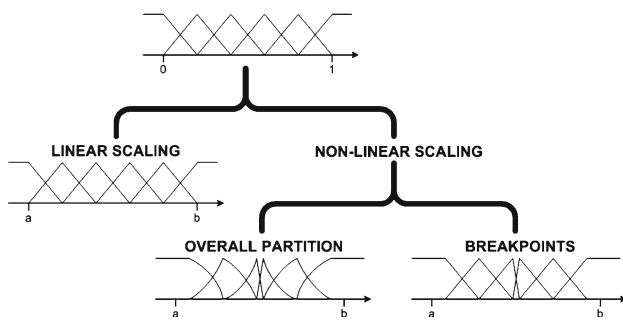
We start from an RB which represents a universal knowledge extracted with an elicitation process by domain experts. Then, adaptation is obtained by applying a modified version of the operators introduced in Botta et al. (2006a).

The paper is organized as follows: Sect. 2 discusses previous work on CA of FRBSs. Section 3 describes the operators used in our CA approach. Section 4 discusses the issue of ordering and interpretability of FRBSs and introduces a novel index to evaluate both these properties concurrently. Section 5 details the MOEA. Section 6 shows two application examples in the fields of regression and data modeling, respectively, and Sect. 7 draws final conclusions.

## 2 Previous work

In the literature, the majority of papers on CA of FRBSs have mostly focused on the use of scaling functions (Bastian 1994; Gudwin and Gomide 1994; Magdalena 1997; Pedrycz et al. 1997; Gudwin et al. 1998; Cordón et al. 2001; Magdalena 2002; Botta et al. 2006a, b). Usually, the scaling function is applied to a normalized partition, that is, a partition defined over the  $[0,1]$  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 function adapts the partition by mapping the normalized universe to the context-adapted universe, possibly modifying the distribution and the shape of fuzzy sets.

The scaling functions used in the literature can roughly be classified into linear (Bastian 1994; Gudwin and Gomide 1994; Magdalena 2002) and non-linear (Magdalena 1997; Pedrycz et al. 1997; Gudwin et al. 1998; Cordón et al. 2001; Klawonn 2006; Botta et al. 2006a, b). Non-linear scaling functions can be applied on the overall universe of discourse, thus modifying the shape of fuzzy sets, as in Magdalena (1997); Pedrycz et al. (1997); Gudwin et al. (1998); Klawonn (2006), or just on some points (e.g., on breakpoints in the case of triangular and trapezoidal fuzzy sets), as in Cordón et al. (2001); Botta et al. (2006ab), so as to maintain the original shape of the fuzzy sets and the interpretability of the partition. Figure 1 shows an example of application of scaling functions to a normalized partition. We note that both linear and non-linear scaling functions comply with the guidelines introduced in Sect. 1. We further remark that scaling functions do not always perform an effective CA, mostly because of their limited modeling capabilities. To overcome this weakness, in Botta et al. (2006b), we introduced four parametric



**Fig. 1** Examples of application of different types of scaling functions to a normalized partition

operators that affect the core, the support and the boundary elements of fuzzy sets. While the new operators improve modeling capabilities of scaling functions, they may also modify the partition in a way that makes the ordering less evident. Furthermore, the interpretability of the resulting partition may be affected as well. In this context, relevant contributions to the identification of parameters of scaling functions and CA operators maximize only the accuracy of the context-adapted FRBS by exploiting evolutionary algorithms (Magdalena 1997; Gudwin et al. 1998; Cordon et al. 2001; Botta et al. 2006a, b).

On the other hand, the issue of balancing interpretability and accuracy has been addressed in the field of FRBS generation from data (de Oliveira 1999; Casillas et al. 2003). Often, this issue has been tackled by using MOEAs aimed at generating a set of FRBSs with different trade-offs between the two objectives (Ishibuchi et al. 1995; Jimenez et al. 2001; Wang et al. 2005; Ishibuchi and Nojima 2007; Gonzalez et al. 2007; Cococcioni et al. 2007). The majority of existing approaches, however, focus on learning the RB from examples rather than on adapting the meaning of linguistic terms to a specific context. Further, interpretability is typically measured indirectly in terms of complexity (for instance, number of fuzzy sets, like in Gonzalez et al. (2007), and/or number of rules) or in terms of properties of the partitions, such as coverage and distinguishability. Finally, when these properties are considered, they are measured by a simple similarity index that cannot completely capture interpretability semantics (Jimenez et al. 2001; Wang et al. 2005).

In this paper, we introduce a novel interpretability 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 (1999). Indeed, 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.

### 3 Operators for CA

In this section, we briefly review the scaling functions and the operators that we use to perform CA of an FRBS. Preliminary versions of these operators have been previously discussed by Botta et al. (2006a). We introduce a flexible and compact scaling function and we define four fuzzy modifiers that allow adapting the core, the support and the boundary elements of fuzzy sets.

In the adaptation process, we first apply the scaling function, and then the fuzzy modifiers. The modifiers are chosen and formulated in such a way that the effects on the context-adapted partition are independent of the order in which they are executed. Thus, modifiers can be applied without interfering with each other. Our modifiers are flexible enough 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 our CA approach, each modifier is applied 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 to be defined next.

The fuzzy modifiers are formulated so as to act on trapezoidal fuzzy sets defined by  $(sl, cl, cu, su)$ , where  $sl$  and  $su$ , and  $cl$  and  $cu$  are the left and right bounds of the support and of the core, respectively, with  $sl \leq cl \leq cu \leq su$ . This definition includes as special cases triangular fuzzy sets (when  $cl = cu$ ) and singletons (when  $sl = cl = cu = su$ ). 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

To cover the overall universe of discourse and to non-uniformly distribute the fuzzy sets in the partition, we adopt the following scaling function:

$$s(x) : [0, 1] \rightarrow [a, b]$$

$$= \begin{cases} a + (b - a)(\lambda^{1-k_{SF}}x^{k_{SF}}) & \text{if } x \leq \lambda \\ a + (b - a)[1 - (1 - \lambda)^{1-k_{SF}}(1 - x)^{k_{SF}}] & \text{if } x > \lambda, \end{cases} \quad (1)$$

where parameters  $a$  and  $b$  identify the bounds of the universe of discourse, parameter  $\lambda \in [0, 1]$  defines a sort of center of gravity in the normalized partition, and parameter  $k_{SF} > 0$  defines the degree of dilation ( $k_{SF} > 1$ ) or compression ( $k_{SF} < 1$ ) of fuzzy sets around  $\lambda$ . For an example of scaling function application, the reader can refer to Botta et al. (2006a).

### 3.2 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 modifier is defined as:

$$cl' = \begin{cases} cl - (sl - cl) \cdot k_{CP} & \text{if } k_{CP} < 0 \\ cl + (su - cu) \cdot k_{CP} & \text{if } k_{CP} \geq 0, \end{cases} \quad (2)$$

$$cu' = \begin{cases} cu - (sl - cl) \cdot k_{CP} & \text{if } k_{CP} < 0 \\ cu + (su - cu) \cdot k_{CP} & \text{if } k_{CP} \geq 0, \end{cases} \quad (3)$$

where  $cl'$  and  $cu'$  are the left and right bounds of the modified core, respectively, and  $k_{CP} \in [-1, 1]$  defines the intensity of the left shift ( $k_{CP} < 0$ ) or right shift ( $k_{CP} > 0$ ) of the core. Figure 2a shows a sample application of the core-position modifier to a partition  $P = \{A_1, \dots, A_5\}$  composed by five uniformly distributed trapezoidal fuzzy sets over the normalized universe  $[0,1]$ .

### 3.3 Core-width modifier

The core-width modifier acts on the core of a fuzzy set, dilating or shrinking the core within the support. The modifier is defined as:

$$cl' = \begin{cases} cl + w \cdot (sl - cl) \cdot k_{CW} & \text{if } k_{CW} < 0 \\ cl + (sl - cl) \cdot k_{CW} & \text{if } k_{CW} \geq 0, \end{cases} \quad (4)$$

$$cu' = \begin{cases} cu + w \cdot (su - cu) \cdot k_{CW} & \text{if } k_{CW} < 0 \\ cu + (su - cu) \cdot k_{CW} & \text{if } k_{CW} \geq 0, \end{cases} \quad (5)$$

where  $w = (cu - cl)/(cl - sl + su - cu)$ , and  $k_{CW} \in [-1, 1]$  determines the intensity of dilation ( $k_{CW} > 0$ ) or shrinking ( $k_{CW} < 0$ ) of the core. Figure 2b shows a sample application of the core-width modifier.

### 3.4 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 modifier is defined as:

$$sl' = sm + k_{SW} \cdot (sl - sm), \quad (6)$$

$$cl' = sm + k_{SW} \cdot (cl - sm), \quad (7)$$

$$cu' = sm + k_{SW} \cdot (cu - sm), \quad (8)$$

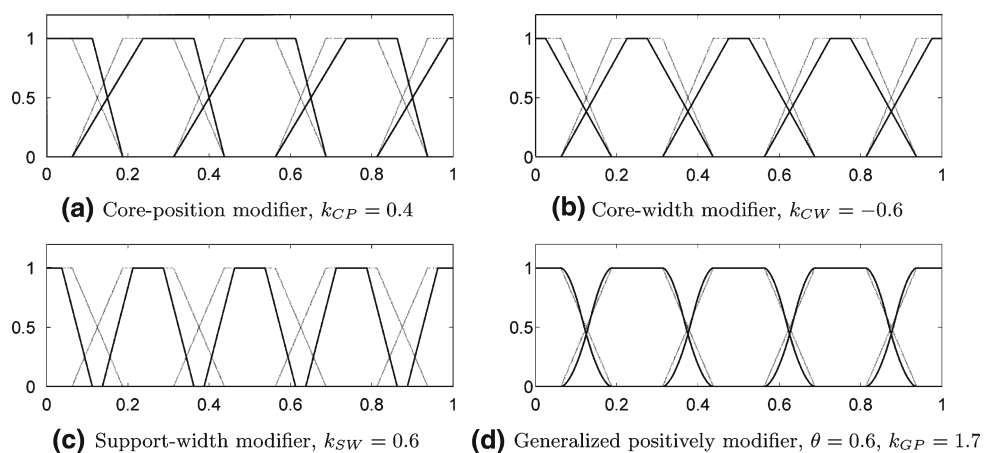
$$su' = sm + k_{SW} \cdot (su - sm), \quad (9)$$

where  $sl'$  and  $su'$  are the left and right bounds of the modified support, respectively,  $sm = (sl + su)/2$ , and  $k_{SW} > 0$  determines the negative ( $k_{SW} < 1$ ) or positive ( $k_{SW} > 1$ ) scaling of the fuzzy set. Figure 2c shows a sample application of the support-width modifier.

### 3.5 Generalized positively modifier

The generalized positively modifier is an extension of the well-known *positively* linguistic hedge. We use this modifier to change the degree of membership of the boundary elements of the fuzzy sets. In other words, the generalized positively modifier changes the shape of the original

**Fig. 2** Sample applications of the modifiers to a partition composed by five trapezoidal fuzzy sets



trapezoidal fuzzy sets so as to generate, for instance, Gaussian-like fuzzy sets. The modifier is defined as:

$$A'(x) = \begin{cases} \theta^{1-k_{GP}} A^{k_{GP}}(x) & \text{if } A(x) < \theta \\ 1 - (1 - \theta)^{1-k_{GP}} [1 - A(x)]^{k_{GP}} & \text{if } A(x) \geq \theta, \end{cases} \tag{10}$$

where  $A'$  is the fuzzy set resulting from the application of the linguistic hedge to  $A$ ,  $\theta \in [0,1]$  is the degree of membership in which the fuzzy set changes concavity, and  $k_{GP} > 0$  is a contrast intensification parameter. Figure 2d shows a sample application of the generalized positively modifier.

### 4 An interpretability index based on fuzzy ordering relations

#### 4.1 Fuzzy sets ordering

As stated in Section 1, CA of an FRBS consists in tuning the fuzzy sets corresponding to the linguistic terms used in the RB. The ordering of the context-adapted fuzzy sets should reflect the semantic ordering of the linguistic terms. The ordering of fuzzy sets has been widely addressed in the literature (Wang and Kerre 2001; Cross and Sudkamp 2002). Given two fuzzy sets  $A_1$  and  $A_2$  defined over the universe of discourse  $\mathbb{R}$ , it is sometimes difficult to determine whether and how  $A_1$  and  $A_2$  are ordered. The 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 sets are clearly and intuitively ordered, but it may lead to counter-intuitive results in the general case. Wang and Kerre (2001) reviewed a number of ordering approaches, including some methods that evaluate the  $\leq$  relation as a fuzzy relation. These methods 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 called *reciprocal*.

Figure 3 shows eight case studies of couples of fuzzy sets (case 4 is taken from an example by Wang and Kerre (2001)). The ordering of cases 1–3 is easily identifiable, while the other cases are disputable. For instance, in case 8,

if we use the interval set ordering of the supports to assess the ordering between  $A_1$  and  $A_2$ , we obtain  $A_2 \leq A_1$ ; on the other hand, if we adopt the ordering of cores, we obtain  $A_1 \leq A_2$ . We used these case studies to benchmark the behavior of Kolodziejczyk’s index and Yuan’s index (Yuan 1991), two reciprocal fuzzy ordering relations previously reviewed by Wang and Kerre (2001). Table 1 shows the values of the two indices for the eight case studies. We remark that, between the two indices, Yuan’s index is able to distinguish one case from another with higher precision.

#### 4.2 Interpretability and ordering of fuzzy partitions

Interpretability of fuzzy partitions can be defined in several ways. Here, we refer to the definition proposed by de Oliveira (1999), who states that a fuzzy partition is interpretable if it satisfies the following properties:

1. The partition should have a reasonable number of fuzzy sets;
2. The fuzzy sets in the partition should all be normal, i.e., for each fuzzy set there exists at least one point with membership degree equal to 1;
3. Each couple of fuzzy sets should be distinguishable enough, so that there are no two fuzzy sets that represent 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 a fuzzy set with a membership degree over a given reasonable threshold.

If the number of linguistic terms is low, as it generally is, guideline 1 defined in Section 1 allows satisfying property 1. Furthermore, if the operators used for CA do not alter the normality of the context-adapted fuzzy sets, as, for instance, the ones introduced in Sect. 3, property 2 is verified as well. On the other hand, properties 3 and 4 are not so easily satisfiable. In Botta et al. (2006a) these properties are achieved by heuristically restricting the domain of parameters of operators. These two properties 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.

An approach to address this problem is to consider distinguishability and coverage as conflicting properties,

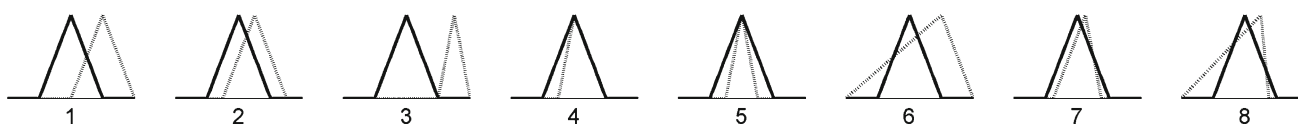


Fig. 3 Case studies for the evaluation of ordering indices between fuzzy sets  $A_1$  (solid) and  $A_2$  (dotted)

**Table 1** Evaluation of  $A_1 \leq A_2$  and  $A_2 \leq A_1$  on the case studies of Fig. 3

	Kolodziejczyk's index		Yuan's index	
	$K \leq (A_1, A_2)$	$K \leq (A_2, A_1)$	$Y \leq (A_1, A_2)$	$Y \leq (A_2, A_1)$
1	1	0	0.9080	0.0920
2	1	0	0.7444	0.2556
3	1	0	1	0
4	1	0	0.5761	0.4239
5	0.5000	0.5000	0.5000	0.5000
6	0.8333	0.1667	0.6576	0.3424
7	0.8333	0.1667	0.5737	0.4263
8	0.4003	0.5997	0.4730	0.5270

and to use a metric to assess the trade-off between them. For instance, the membership value of crossing points between adjacent fuzzy sets can be used to measure both at a time (de Oliveira 1999). 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. For instance, by using the membership degrees of crossing points of adjacent fuzzy sets, we can only derive a partial ordering based on  $\alpha$ -cuts.

Further, it can be shown that Jaccard's index (Cross and Sudkamp 2002), which is often used to measure interpretability, does not seem to provide means for salient evaluation of ordering of fuzzy partitions. Some ordering indices, however, can be used to evaluate distinguishability and coverage from an interpretability point-of-view.

Indeed, let us consider again Fig. 3 and Table 1. 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 property and case 4 is lacking distinguishability property. The 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, the 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 is a sensible choice for evaluating interpretability.

### 4.3 The interpretability index

Let us consider a partition  $P = \{A_1, \dots, A_i, \dots, A_N\}$  consisting of  $N$  fuzzy sets. Let  $d_{j,i} = |j - i|$  be the semantic distance between  $A_j$  and  $A_i$ . For instance, the semantic distance

between  $A_3$  and  $A_1$  is 2. We define the following index to evaluate interpretability and ordering of a partition:

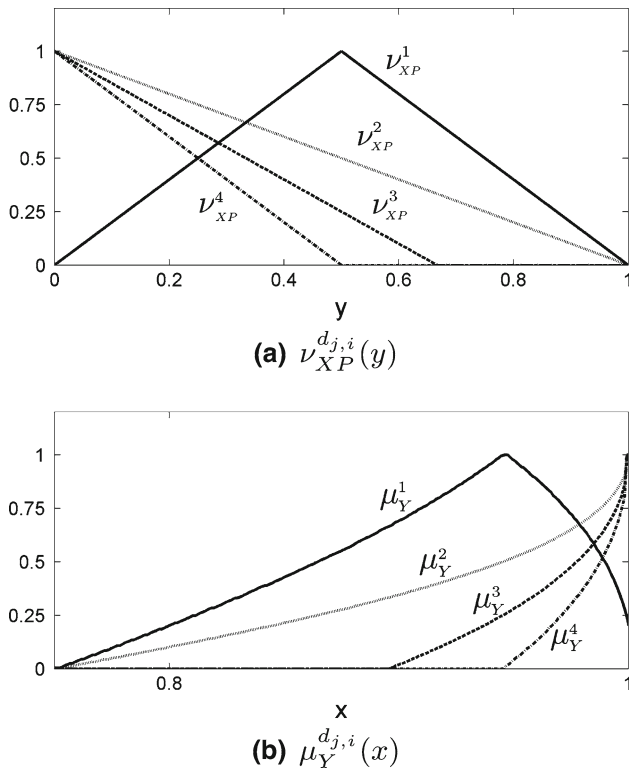
$$\Phi_Q(P) = \frac{\sum_{1 \leq i \leq N-1} \sum_{i < j \leq N} \frac{1}{d_{j,i}} \cdot \mu_Q^{d_{j,i}}(Q \leq (A_i, A_j))}{\sum_{1 \leq i \leq N-1} \sum_{i < j \leq N} \frac{1}{d_{j,i}}}, \tag{11}$$

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

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  $\mu_Q^{d_{j,i}}(x)$  are used to assess, for different values of  $d_{j,i}$ , the value of the  $Q$  index with respect to interpretability. For instance, in case of  $d_{j,i} = 1$  and  $A_i \cap A_j = \emptyset$ , i.e., a situation similar to case 3 in Fig. 3, adjacent fuzzy sets are not overlapped and, therefore, coverage is not verified. It follows that  $\mu_Q^{d_{j,i}}(x)$  should return a value close to 0. On the other hand, in case of  $d_{j,i} > 1$  and  $A_i \cap A_j = \emptyset$ , to enforce distinguishability, the fuzzy sets should not be overlapped and, therefore,  $\mu_Q^{d_{j,i}}(x)$  should return a value close to 1. Obviously, the definition of the family of fuzzy sets  $\mu_Q^{d_{j,i}}(x)$  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  $\mu_Q^{d_{j,i}}(x)$ .

We start from the evaluation of a well-known measure of interpretability, based on the value  $y$  of the crossing point between two fuzzy sets (denoted as XP in the following). We can distinguish two cases:

- $d_{j,i} = 1$  (*Semantically adjacent fuzzy sets*): 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  $v_{XP}^1(y)$  shown in solid line in Fig. 4a;
- $d_{j,i} > 1$  (*Semantically non-adjacent fuzzy sets*): 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  $d_{j,i}$ , we can still tolerate some overlapping between the two fuzzy sets, and this tolerance should decrease with the increase of  $d_{j,i}$ . These observations can be modeled by the fuzzy set  $v_{XP}^{d_{j,i}}(y)$  shown in dotted line in Fig. 4a for different values of  $d_{j,i}$ . We note that, while  $v_{XP}^{d_{j,i}}(0) = 1 \forall d_{j,i} > 1$ , the right spread shrinks toward 0 as  $d_{j,i}$  increases, so as to reduce the tolerance.



**Fig. 4** Fuzzy sets  $v_{XP}^{d_{j,i}}(y)$  and  $\mu_Y^{d_{j,i}}(x)$  used in the examples of Sect. 6

To obtain  $\mu_Q^{d_{j,i}}(x)$ , we project  $v_{XP}^{d_{j,i}}(y)$  from its original universe of discourse  $y$  to the universe of discourse  $x$ . The overall process can be formalized as follows:

1. We choose  $v_{XP}^{d_{j,i}}(y)$  as triangular membership functions, defined by the three breakpoints  $(sl^{d_{j,i}}, c^{d_{j,i}}, su^{d_{j,i}})$ . We set  $sl^{d_{j,i}} = 0 \forall d_{j,i}, c^1 = 0.5, c^{d_{j,i}} = 0 \forall d_{j,i} > 1, su^1 = 1$ , and  $su^{d_{j,i}} = 2/d_{j,i} \forall d_{j,i} > 1$ . Figure 4a shows  $v_{XP}^{d_{j,i}}(y)$ , with  $N = 5$  and  $d_{j,i} = 1..4$ ;
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  $v_{XP}^{d_{j,i}}(y)$  from the universe of discourse  $y$  to the universe of discourse  $x$ , thus obtaining a corresponding  $\mu_Q^{d_{j,i}}(x)$ . More formally, we have:

$$\mu_Q^{d_{j,i}}(x) = \sup_{y \in [0,1]} \min(v_{XP}^{d_{j,i}}(y), R_{Q,XP}(x,y)), \tag{12}$$

where  $d_{j,i} = 1..N-1$ .

We illustrate the above process by detailing the sets  $\mu_Y^{d_{j,i}}(x)$  corresponding to Yuan’s fuzzy ordering index. We recall the formula to compute Yuan’s index (Yuan 1991; Wang and Kerre 2001):

$$Y \leq (A_1, A_2) = \Delta_{A_2, A_1} / (\Delta_{A_1, A_2} + \Delta_{A_2, A_1}), \tag{13}$$

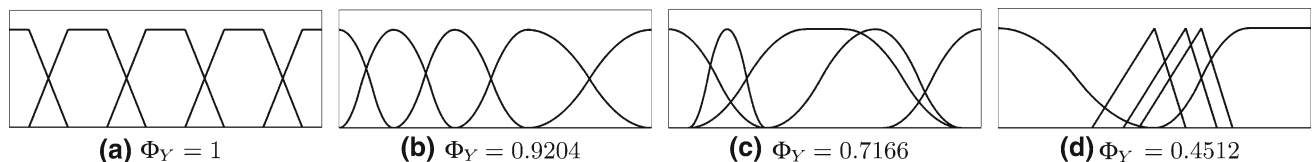
where  $\Delta_{A_1, A_2}$  is defined as:

$$\Delta_{A_1, A_2} = \int_{\alpha | u_{A_{1z}} > l_{A_{2z}}} (u_{A_{1z}} - l_{A_{2z}}) d\alpha + \int_{\alpha | l_{A_{1z}} > u_{A_{2z}}} (l_{A_{1z}} - u_{A_{2z}}) d\alpha,$$

$\alpha \in [0,1]$  is an  $\alpha$ -cut value,  $A_i \alpha$  is the crisp set obtained by  $\alpha$ -cutting  $A_i, i = \{1,2\}$ , and  $l_{A_{iz}}$  and  $u_{A_{iz}}$  are the lower and upper bounds of  $A_{iz}$ , respectively. Figure 4b shows the projections  $\mu_Y^{d_{j,i}}(x)$  obtained by applying the overall process. This family of fuzzy sets can then be employed in the corresponding interpretability index  $\Phi_Y(P)$ . In Fig. 5, we show the evaluation of the proposed index on four sample partitions characterized by different degrees of coverage, distinguishability and ordering of fuzzy sets.

### 5 The MOEA

Thanks to their modeling capabilities, the scaling function and the fuzzy modifiers introduced in Sect. 3 allow adapting the normalized partitions to any context. In particular, we apply a MOEA to perform the CA described in Sect. 1. To this aim, we need to determine which operators should be applied and the values of their parameters. These choices are typically based on the maximization of the accuracy of the FRBS on real-world examples which describe the behavior of the context-adapted system. Furthermore, as stated in Sects. 1 and 4, the partition generated by the application of the operators should satisfy the ordering and interpretability constraints. Our objectives are accuracy, measured by MSE, and interpretability, measured by the average  $\bar{\Phi}_Y$  of the values of the index  $\Phi_Y$  computed for all input and output partitions of the FRBS. These conflicting goals are balanced by the adoption of NSGA-II (Deb et al. 2002).



**Fig. 5** Evaluation of  $\Phi_Y$  on sample partitions with different degrees of coverage, distinguishability and ordering of fuzzy sets

NSGA-II is a fast and elitist MOEA based on a non-dominance rank assignment and an ad-hoc density-estimation metric. In our CA approach based on scaling functions and fuzzy modifiers, individuals represent the values of the parameters which define the operators used to adapt a Mamdani-type FRBS.

Assume that the FRBS has  $V-1$  input variables and one output. Let  $S = (\hat{x}_j, \hat{y}_j)$  be a set of  $M$  real-world examples, where  $\hat{x}_j$  is a vector of  $V-1$  input values, and  $\hat{y}_j$  is the corresponding output. Each individual in the population is represented by a chromosome composed of  $V$  strings of 77 bits, where each string codes the five control genes and the nine 8-bit parameters of the tuning operators (the first  $V-1$  strings determine the parameters for the input variables and the last string for the output variable). Each string is characterized by a hierarchical structure: the first five bits, one for each operator, control whether the corresponding operator is applied or not on each fuzzy partition. The other 72 bits are organized in sub-strings of 8 bits: each sub-string determines the value of a different parameter.

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 exploring 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. To solve this problem, we have adopted, as usual when using binary chromosomes, the Gray decoding to generate individuals from chromosomes (Michalewicz 1999). The chromosome has the following structure:

$$\begin{aligned} &(C_{SF}^1, C_{CP}^1, C_{CW}^1, C_{SW}^1, C_{GP}^1, \\ &a^1, b^1, \lambda^1, k_{SF}^1, k_{CP}^1, k_{CW}^1, k_{SW}^1, \theta^1, k_{GP}^1, \\ &\dots \\ &C_{SF}^V, C_{CP}^V, C_{CW}^V, C_{SW}^V, C_{GP}^V, \\ &a^V, b^V, \lambda^V, k_{SF}^V, k_{CP}^V, k_{CW}^V, k_{SW}^V, \theta^V, k_{GP}^V), \end{aligned}$$

where  $C_{SF}^v, C_{CP}^v, C_{CW}^v, C_{SW}^v$ , and  $C_{GP}^v$ , with  $v = 1, \dots, V$ , are the control genes that determine whether, respectively, the non-linear scaling function of Eq. 1, the core-position modifier, the core-width modifier, the support-width modifier and the generalized positively modifier have to be applied to the partition. The  $a^v, \dots, k_{SF}^v$  and  $k_{CP}^v, \dots, k_{GP}^v$  are, respectively, the values of the parameters of the scaling function and of the four fuzzy modifiers for the  $v$ th partition.

To the aim of covering the overall universe of discourse, even if  $C_{SF}^v = 0$ , we perform a linear scaling from the normalized universe  $[0,1]$ , in which the initial uniform partition is defined, to the context-adapted universe  $[a^v, b^v]$ . 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 might 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.

We start with an initial population composed of randomly generated individuals. At each generation, the uniform crossover and the uniform mutation operators are applied (Michalewicz 1999), with probability of 0.8 and 0.05, respectively. Chromosomes to be mated are chosen by the standard binary tournament proposed by Deb et al. (2002) in the original version of NSGA-II.

## 6 Experimental results

To evaluate the effectiveness of our approach, we applied the proposed CA method to a regression problem and a data modeling problem. In the first experiment, contexts are modeled by similar curves generated from a parametric function. In the second experiment, we determine the fuel efficiency of a set of vehicles in the contexts of city and highway traffic conditions.

### 6.1 Parametric function

Let us consider the following parametric function:

$$g(x_1, x_2) = \kappa + e^{-(\kappa \cdot x_1)^{2\kappa} - (1+x_2)^{2\kappa}} - e^{-x_1^{2\kappa} - x_2^{2\kappa}} - e^{-(1+x_1)^{2\kappa} - (\kappa \cdot x_2)^{2\kappa}}, \quad (14)$$

where  $x_1, x_2 \in [-1.5, 0.5]$ . We evaluated the function with  $\kappa$  in  $\{2, 5, 7\}$ . The range of the output variable and the smoothness of each curve display similar characteristics for different values of parameter  $\kappa$ . Hence, we can define a set of rules that linguistically describe these common features, and consider each of the three curves as a different instantiation of the same generic shape in a different context, determined by the value of  $\kappa$ . We use the intuitive set of rules shown in Table 2. In the rule base, four linguistic labels are used in each universe, namely *L* (low), *ML* (medium-low), *MH* (medium-high) and *H* (high). For each of the three instances, we applied the CA approach introduced in the previous section. We evaluated the curves in a



**Table 2** RB for the parametric function data set

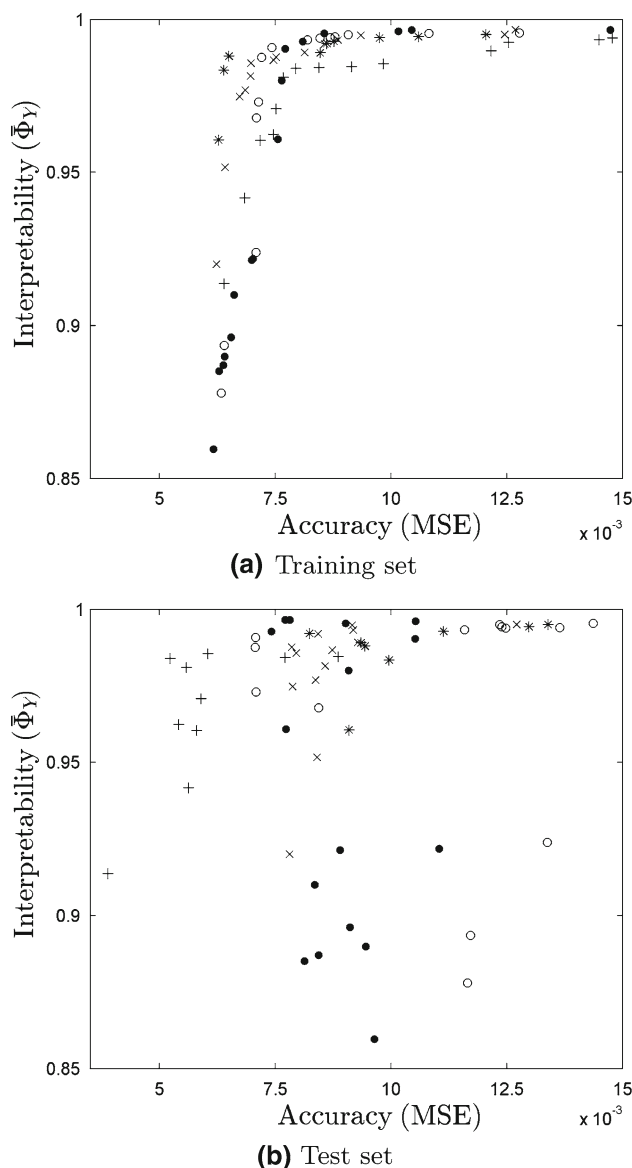
		$x_1$			
		$L$	$ML$	$MH$	$H$
$x_2$	$H$	ML	L	L	ML
	$MH$	ML	L	ML	ML
	$ML$	MH	ML	MH	MH
	$L$	MH	MH	H	H

grid of 120 equally spaced points chosen in the  $[-1.5, 0.5] \times [-1.5, 0.5]$  region. Since there are four linguistic labels for each variable in the RB, the normalized universes of discourse were uniformly partitioned into four trapezoidal fuzzy sets. To evaluate the robustness of MOEA in different runs, we adopted a fivefold cross-validation. Figure 6 shows the Pareto fronts obtained by MOEA 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 quite close to each other on the training set, thus highlighting that the fronts do not depend on the particular execution of MOEA. Due to lack of space, we do not show the Pareto fronts obtained for the other contexts. However, we remark that similar trends can be observed in the other contexts.

To assess the goodness of the Pareto fronts, we have also applied the following CA methods as comparison:

- The non-linear scaling function proposed by Cordón et al. (2001), optimizing its parameters by a single-objective genetic algorithm (SOGA) with binary representation of parameters in the chromosome, uniform mutation, uniform crossover and binary tournament selection. We refer to this approach as SOGA<sub>1</sub>. We remark that we use only the scaling function and not the overall methodology proposed by Cordón et al. (2001) 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.
- The non-linear scaling function and the four fuzzy modifiers introduced in Section 3, optimizing their parameters by the same SOGA as in the previous item. We refer to this method as SOGA<sub>2</sub>.

In both SOGAs, we adopted the same crossover and mutation probabilities as in NSGA-II. Further, to guarantee a fair comparison, we used the same chromosome coding and genetic operators for all the approaches. Finally, we employed the MSE as the single-objective fitness function. All the approaches were tested with a population of  $N_{pop} = 50$  individuals and a maximum number of generations  $I_{max} = 200$ . The experiments on SOGA<sub>1</sub> and SOGA<sub>2</sub> were repeated five times for each context and for each fold to obtain statistically meaningful results.



**Fig. 6** Pareto fronts obtained on the training set (*top*) and on the test set (*bottom*) for the parametric function data set on context  $\kappa = 2$

Table 3 shows the comparison among the results, expressed as average  $\pm$  standard deviation, achieved by the three techniques considered. For MOEA, we selected two FRBSs: one, denoted as MOEA<sub>a</sub>, with the lowest MSE on the test set, and the other, denoted as MOEA<sub>b</sub>, with the lowest MSE among the solutions dominating the FRBSs determined by SOGA<sub>1</sub> on the test set.

As expected, on the test set the context-adapted FRBSs generated by SOGA<sub>2</sub> are characterized by a low MSE, comparable, however, to the best MSEs obtained by the FRBSs in the Pareto fronts. Nevertheless, their interpretability is poor, since the only objective of SOGA<sub>2</sub> is to minimize the MSE. SOGA<sub>1</sub> achieves values of MSE higher than SOGA<sub>2</sub>, but generates more interpretable FRBSs,

**Table 3** Results for the parametric function data set

	SOGA <sub>1</sub>	SOGA <sub>2</sub>	MOEA <sub>a</sub>	MOEA <sub>b</sub>
$\kappa = 2$				
MSE <sub>train</sub>	0.0073 ± 0.0012	0.0056 ± 0.0004	0.0075 ± 0.0007	0.0075 ± 0.0007
MSE <sub>test</sub>	0.0094 ± 0.0032	0.0068 ± 0.0016	0.0066 ± 0.0017	0.0066 ± 0.0017
$\bar{\Phi}_Y$	0.8250 ± 0.0155	0.7003 ± 0.0350	0.9481 ± 0.0443	0.9481 ± 0.0007
$\kappa = 5$				
MSE <sub>train</sub>	0.0462 ± 0.0036	0.0441 ± 0.0034	0.0526 ± 0.0053	0.0526 ± 0.0053
MSE <sub>test</sub>	0.0575 ± 0.0152	0.0554 ± 0.0111	0.0509 ± 0.0139	0.0509 ± 0.0139
$\bar{\Phi}_Y$	0.8510 ± 0.0084	0.7403 ± 0.0484	0.9921 ± 0.0024	0.9921 ± 0.0024
$\kappa = 7$				
MSE <sub>train</sub>	0.0777 ± 0.0056	0.0684 ± 0.0068	0.0817 ± 0.0125	0.0836 ± 0.0106
MSE <sub>test</sub>	0.0918 ± 0.0250	0.0842 ± 0.0273	0.0811 ± 0.0333	0.0812 ± 0.0334
$\bar{\Phi}_Y$	0.8899 ± 0.0156	0.6307 ± 0.0106	0.9533 ± 0.0750	0.9880 ± 0.0115

since it is a scaling function-based approach and, therefore, introduces a lower distortion in the fuzzy partitions than SOGA<sub>2</sub>.

MOEA provides the decision maker with a set of FRBSs with different trade-offs between accuracy and interpretability. In particular, MOEA<sub>a</sub> and MOEA<sub>b</sub> achieve an MSE equal to or lower than SOGA<sub>1</sub> and SOGA<sub>2</sub> on the test set, and are characterized by higher values of  $\bar{\Phi}_Y$ . Although SOGA<sub>1</sub> performs better than MOEA on the training set, MOEA outperforms it on test set. Indeed, due to its simplicity, SOGA<sub>1</sub> can perform a deeper exploration of the search space than MOEA, but can also easily incur overfitting problems. We recall that we used the same number of generations for both SOGAs and MOEA. On the other hand, MOEA balances MSE with interpretability and, therefore, is less prone to overfitting than SOGA<sub>1</sub>. Further, we observe that, as expected, solutions generated by SOGA<sub>2</sub> actually lie on a hypothetical extension of the Pareto front in a zone of low interpretability.

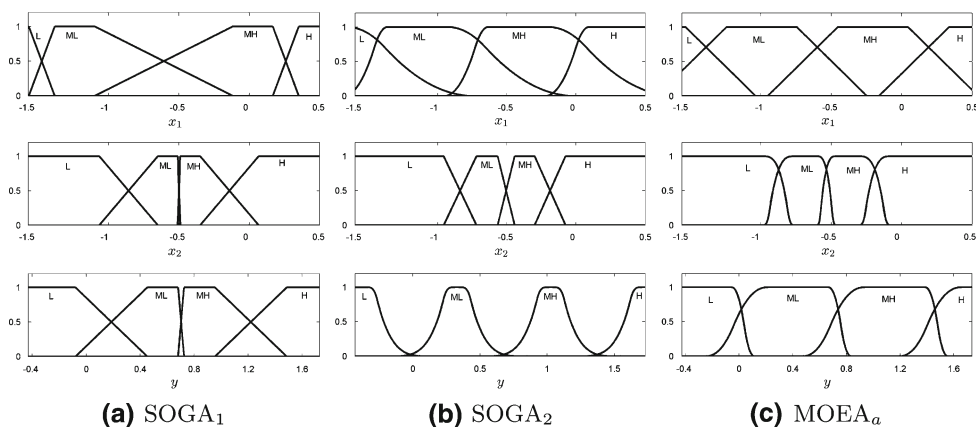
Figure 7 shows, for the context corresponding to  $\kappa = 2$ , an example of the fuzzy partitions of the input and output

variables of the FRBSs generated by SOGA<sub>1</sub> (Fig. 7a) and SOGA<sub>2</sub> (Fig. 7b), and the fuzzy partitions of MOEA<sub>a</sub> which, in this context, corresponds also to MOEA<sub>b</sub> (Fig. 7c). We note that the FRBS generated by SOGA<sub>2</sub> lacks coverage on  $y$ , whereas MOEA<sub>a</sub> shows a high interpretability degree and a low MSE.

### 6.2 Fuel efficiency

The 2004 new car and truck data set (Johnson 2004) 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 computing the correlation between FE and the other features, 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

**Fig. 7** Partitions generated for the parametric function data set on context  $\kappa = 2$



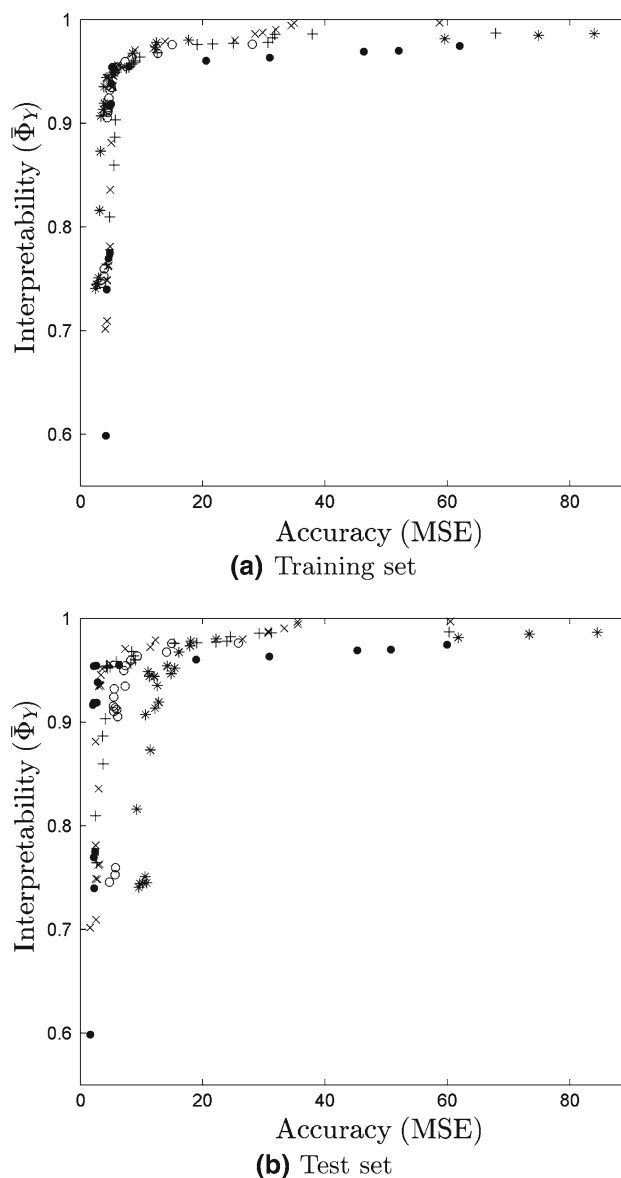
**Table 4** RB for the fuel efficiency data set

Rule	ES	HP	AW	RP	FE
$R_1$	L	L	M	L	H
$R_2$	L	L	H	L	H
$R_3$	L	M	M	L	M
$R_4$	L	M	H	L	H
$R_5$	M	M	L	M	M
$R_6$	M	M	L	H	M
$R_7$	M	M	M	M	M
$R_8$	M	M	H	M	M
$R_9$	M	H	L	M	M
$R_{10}$	M	H	L	H	L
$R_{11}$	M	H	M	M	M
$R_{12}$	M	H	M	H	L
$R_{13}$	M	H	H	M	M
$R_{14}$	H	M	L	M	M
$R_{15}$	H	M	L	H	L
$R_{16}$	H	M	M	M	M
$R_{17}$	H	M	M	H	L
$R_{18}$	H	H	L	M	L
$R_{19}$	H	H	L	H	L
$R_{20}$	H	H	M	M	M
$R_{21}$	H	H	M	H	L

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. City and highway traffic conditions were considered as two different contexts. ES, HP, AW, RP and FE are measured in liters, horsepower, inches<sup>2</sup>/pounds, dollars and miles per gallon (MPG), respectively.

The rules of the FRBS, shown in Table 4, were extracted from the following intuitive considerations derived from experience: 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*. Thus, we uniformly partitioned the normalized input variables and the output variable into three fuzzy sets, namely *L (low)*, *M (medium)*, and *H (high)*. The number of fuzzy sets was chosen by interviewing a pool of experts and asking them for a meaningful partition of the universes. Again, a fivefold cross-validation was performed, with  $N_{pop} = 50$  and  $I_{max} = 200$ . Figure 8 shows the Pareto fronts obtained for each fold and for both the training and test sets on the city context (in the highway context, we observed similar fronts). Table 5 summarizes the results obtained by comparing MOEA, SOGA<sub>1</sub> and SOGA<sub>2</sub> as described in Sect. 6.1.

Figure 8 and Table 5 confirm the trend that we observed on the other data set. Indeed, the FRBSs determined by



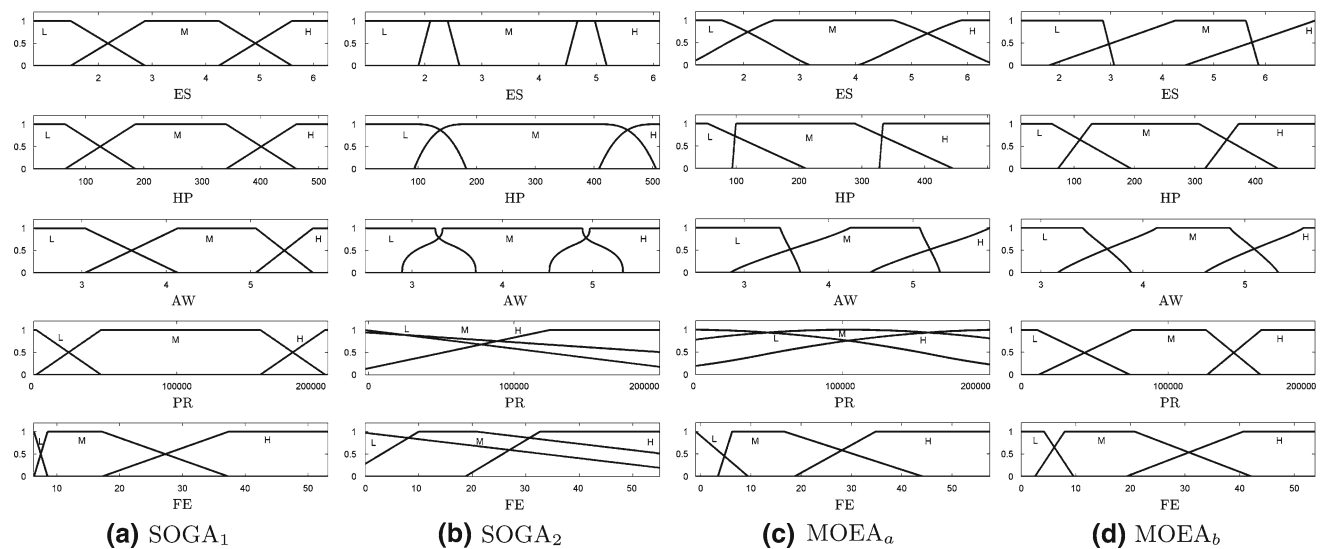
**Fig. 8** Pareto fronts obtained on the training set (*top*) and on the test set (*bottom*) for the fuel efficiency data set in the city context

SOGA<sub>2</sub> are characterized by a low MSE and a poor interpretability, while SOGA<sub>1</sub> generates FRBSs with good trade-offs between accuracy and interpretability that are, however, Pareto-dominated by some solutions found by our MOEA. Further, on this dataset, the FRBSs identified by NSGA-II generalize much better than FRBSs generated by SOGA<sub>1</sub> and SOGA<sub>2</sub>, since they achieve similar MSEs 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<sub>1</sub>.

Figure 9 shows, for the city context, an example of the fuzzy partitions of the input and output variables of the FRBSs chosen as in Sect. 6.1. We note that the partitions of

**Table 5** Results for the fuel efficiency data set

	SOGA <sub>1</sub>	SOGA <sub>2</sub>	MOEA <sub>a</sub>	MOEA <sub>b</sub>
<b>City context</b>				
MSE <sub>train</sub>	5.0259 ± 0.4475	3.0093 ± 0.5967	3.6916 ± 0.7755	4.6980 ± 0.8706
MSE <sub>test</sub>	5.4802 ± 3.0011	3.8554 ± 2.7852	3.6821 ± 2.8287	4.7619 ± 3.5105
$\bar{\Phi}_Y$	0.8525 ± 0.0120	0.4736 ± 0.0640	0.7273 ± 0.0664	0.8987 ± 0.0173
<b>Highway context</b>				
MSE <sub>train</sub>	6.8556 ± 0.9315	3.8348 ± 0.3534	4.1016 ± 0.3193	6.6064 ± 1.0363
MSE <sub>test</sub>	7.6167 ± 2.0212	4.6913 ± 1.7287	4.6426 ± 1.9782	6.3652 ± 2.4641
$\bar{\Phi}_Y$	0.8731 ± 0.0108	0.4778 ± 0.0940	0.7740 ± 0.0402	0.9229 ± 0.0365

**Fig. 9** Partitions generated for the fuel efficiency data set on the city context

the FRBS generated by SOGA<sub>2</sub>, 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, MOEA<sub>a</sub> achieves the lowest MSE on the test set and it maintains its interpretability, even if the distinguishability of fuzzy sets on the PR input is not completely evident. Finally, MOEA<sub>b</sub> (Fig. 9d) outperforms all the other FRBSs in terms of interpretability and, thanks to the modeling power of fuzzy modifiers, achieves an MSE lower than SOGA<sub>1</sub>.

## 7 Conclusion

In this paper, we have proposed a MOEA-based approach to context adaptation of Mamdani-type FRBSs. MOEA generates a Pareto front composed of context-adapted FRBSs with different trade-offs between accuracy and interpretability. Accuracy is measured in terms of mean

square error. Interpretability is evaluated using a novel index based on a fuzzy ordering relation. This index can be employed in any constrained tuning of an FRBS where ordering and interpretability of fuzzy partitions are required. The results obtained by our approach using a Yuan-based interpretability index on synthetic and real data sets have shown that the MOEA-based CA can determine solutions that achieve an error equal to or better than accuracy-oriented approaches, while, at the same time, preserving interpretability of the fuzzy partitions.

## References

- Bastian A (1994) How to handle the flexibility of linguistic variables with applications. *Int J Uncert Fuzziness Knowl Based Syst* 2(4):463–484
- Botta A, Lazzerini B, Marcelloni F (2006a) Context adaptation of Mamdani fuzzy systems through new operators tuned by a genetic algorithm. In: *Proceedings of 2006 IEEE international conference on fuzzy systems*. Vancouver, BC, pp 1641–1648

- Botta A, Lazzerini B, Marcelloni F (2006b) New operators for context adaptation of Mamdani fuzzy systems. In: Ruan D, D'hondt P, Fantoni PF, Cock MD, Nachtegaele M, Kerre EE (eds) Applied artificial intelligence. World Scientific, Singapore, pp 35–42
- Casillas J, Cordon O, Herrera F, Magdalena L (eds) (2003) Interpretability issues in fuzzy modeling, vol 128. Springer, Heidelberg
- Cococcioni M, Ducange P, Lazzerini B, Marcelloni F (2007) A Pareto-based multi-objective evolutionary approach to the identification of Mamdani fuzzy systems. *Soft Comput* 11(11):1013–1031
- Cordon O, Herrera F, Magdalena L, Villar P (2001) A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule-based system data base. *Inform Sci* 136(1–4): 85–107
- Cordon O, Gomide F, Herrera F, Hoffmann F, Magdalena L (2004) Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets Syst* 141:5–31
- Cross VV, Sudkamp TA (2002) Similarity and compatibility in fuzzy set theory. Springer, Heidelberg
- Deb K, Agrawal S, Pratap A, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
- Gonzalez J, Rojas I, Pomares H, Herrera LJ, Guillen A, Palomares JM, Rojas F (2007) Improving the accuracy while preserving the interpretability of fuzzy function approximators by means of multi-objective evolutionary algorithms. *Int J Approx Reason* 44(1):32–44
- Gudwin R, Gomide F, Pedrycz W (1998) Context adaptation in fuzzy processing and genetic algorithms. *Int J Intell Syst* 13(10–11):929–948
- Gudwin RR, Gomide FAC (1994) Context adaptation in fuzzy processing. In: Proceedings of Brazil–Japan joint symposium on fuzzy systems, Campinas, Brazil
- Ishibuchi H, Nojima Y (2007) Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *Int J Approx Reason* 44(1):4–31
- Ishibuchi H, Murata T, Turksen IB (1995) Selecting linguistic classification rules by two-objective genetic algorithms. In: Proceedings of IEEE international conference on systems, man and cybernetics, vol 2. Vancouver, BC, pp 1410–1415
- Jimenez F, Gomez-Skarmeta AF, Roubos H, Babuska R (2001) Accurate, transparent, and compact fuzzy models for function approximation and dynamic modeling through multi-objective evolutionary optimization. In: Proceedings of 1st international conference on evolutionary multi-criterion optimization, Zurich, Switzerland, vol 2, pp 653–667
- Johnson RW (2004) New car and truck data. <http://www.amstat.org/publications/jse>
- Klawonn F (2006) Reducing the number of parameters of a fuzzy system using scaling functions. *Soft Comput* 10(9):749–756
- Magdalena L (1997) Adapting the gain of an FLC with genetic algorithms. *Int J Approx Reason* 17(4):327–349
- Magdalena L (2002) On the role of context in hierarchical fuzzy controllers. *Int J Intell Syst* 17(5):471–493
- Mamdani E, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man Machine Stud* 7(1):1–13
- Michalewicz Z (1999) Genetic algorithms + data structures = evolution programs. Springer, Heidelberg
- de Oliveira JV (1999) Semantic constraints for membership function optimization. *IEEE Trans Syst Man Cybern Part A* 29(1):128–138
- Pedrycz W, Gudwin RR, Gomide FAC (1997) Nonlinear context adaptation in the calibration of fuzzy sets. *Fuzzy Sets Syst* 88(1):91–97
- Wang H, Kwong S, Jin Y, Wei W, Man KF (2005) Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. *Fuzzy Sets Syst* 149(1):149–186
- Wang X, Kerre EE (2001) Reasonable properties for the ordering of fuzzy quantities. *Fuzzy Sets Syst* 118(3):375–405
- Yuan Y (1991) Criteria for evaluating fuzzy ranking methods. *Fuzzy Sets Syst* 43(2):139–157