Interpretability Improvements to Find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview

Jorge Casillas[1], Oscar Cordón[1], Francisco Herrera[1], and Luis Magdalena[2]

[1] Department of Computer Science and Artificial Intelligence,
University of Granada, E-18071 Granada, Spain
e-mail: {casillas,ocordon,herrera}@decsai.ugr.es
[2] Department of Mathematics Applied to Information Technologies,
Technical University of Madrid, E-28040 Madrid, Spain
e-mail: llayos@mat.upm.es

Abstract. System modeling with fuzzy rule-based systems (FRBSs), i.e. fuzzy modeling (FM), usually comes with two contradictory requirements in the obtained model: the *interpretability*, capability to express the behavior of the real system in an understandable way, and the *accuracy*, capability to faithfully represent the real system. While linguistic FM (mainly developed by linguistic FRBSs) is focused on the interpretability, precise FM (mainly developed by Takagi-Sugeno-Kang FRBSs) is focused on the accuracy. Since both criteria are of vital importance in system modeling, the balance between them has started to pay attention in the fuzzy community in the last few years.

The chapter analyzes mechanisms to find this balance by improving the interpretability in linguistic FM: selecting input variables, reducing the fuzzy rule set, using more descriptive expressions, or performing linguistic approximation; and in precise FM: reducing the fuzzy rule set, reducing the number of fuzzy sets, or exploiting the local description of the rules.

1 Introduction

System modeling is the action and effect of approaching to a model, i.e., to a theoretical scheme that simplifies a real system or complex reality with the aim of easing its understanding. Thanks to these models, the real system can be explained, controlled, simulated, predicted, and even improved. The development of *reliable* and *comprehensible* models is the main objective in system modeling. If not so, the model loses its usefulness.

There are at least three different paradigms in system modeling. The most traditional approach is the *white box modeling*, which assumes that a thorough knowledge of the system's nature and a suitable mathematical scheme to represent it are available. As opposed to it, the *black box modeling* [60] is performed entirely from data using no additional a priori knowledge and considering a sufficiently general structure. Whereas the white box modeling has serious difficulties when complex and poorly understood systems are considered, the black box modeling deals with structures and associated parameters

that usually do not have any physical significance [2]. Therefore, generally the former approach does not adequately obtain reliable models while the latter one does not adequately obtain comprehensible models.

A third, intermediate approach arises as a combination of the said paradigms, the *grey box modeling* [28], where certain known parts of the system are modeled considering the prior understood and the unknown or less certain parts are identified with black box procedures. With this approach, the mentioned disadvantages are palliated and a better balance between reliability and comprehensibility is attained.

Nowadays, one of the most successful tools to develop grey box models is *fuzzy modeling* (FM) [41], which is an approach used to model a system making use of a descriptive language based on fuzzy logic with fuzzy predicates [63]. FM usually considers model structures (fuzzy systems) in the form of fuzzy rule-based systems (FRBSs) and constructs them by means of different parametric system identification techniques. Fuzzy systems have demonstrated their ability for control [17], modeling [49], or classification [12] in a huge number of applications. The keys for their success and interest are the ability to incorporate human expert knowledge – which is the information mostly provided for many real-world systems and is described by vague and imprecise statements – and the facility to express the behavior of the system with a language easily interpretable by human beings. These interesting advantages allow them to be even used as mechanisms to interpret black box models such as neural networks [11].

As a system modeling discipline, FM is mainly characterized by two features that assess the quality of the obtained fuzzy models:

- *Interpretability* — It refers to the capability of the fuzzy model to express the behavior of the system in a understandable way. This is a subjective property that depends on several factors, mainly the model structure, the number of input variables, the number of fuzzy rules, the number of linguistic terms, and the shape of the fuzzy sets. With the term interpretability we englobe different criteria appeared in the literature such as *compactness*, *completeness*, *consistency*, or *transparency*.
- *Accuracy* — It refers to the capability of the fuzzy model to faithfully represent the modeled system. The closer the model to the system, the higher its accuracy. As closeness we understand the similarity between the responses of the real system and the fuzzy model. This is why the term approximation is also used to express the accuracy, being a fuzzy model a fuzzy function approximation model.

As Zadeh stated in its *Principle of Incompatibility* [75], "*as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics.*"

Therefore, to obtain high degrees of interpretability and accuracy is a contradictory purpose and, in practice, one of the two properties prevails over the other one. Depending on what requirement is mainly pursued, the FM field may be divided into two different areas:

- *Linguistic fuzzy modeling (LFM)* — The main objective is to obtain fuzzy models with a good interpretability.
- *Precise fuzzy modeling (PFM)* — The main objective is to obtain fuzzy models with a good accuracy.

The relatively easy design of fuzzy systems, their attractive advantages, and their emergent proliferation have made FM to suffer a deviation from the seminal purpose directed towards exploiting the descriptive power of the concept of a linguistic variable [75,76]. Instead, in the last few years, the prevailing research in FM has focused on increasing the accuracy as much as possible paying little attention to the interpretability of the final model.

Nevertheless, a new tendency in the FM scientific community that looks for a good balance between interpretability and accuracy is increasing in importance [3,9,54,65]. The aim of this chapter is to review some of the recent proposals that attempt to address this issue using mechanisms to improve the interpretability of fuzzy models.

The chapter is organized as follows. Section 2 analyzes the different existing lines of research related to the improvement of interpretability and accuracy to find a good balance in FM, Sect. 3 introduces the most important kinds of FRBSs used to improve their interpretability, Sect. 4 shows how to improve the interpretability of linguistic fuzzy models, Sect. 5 introduces tools to improve the interpretability of precise fuzzy models and, finally, Sect. 6 points out some conclusions.

2 Major Lines of Work

The two main objectives to be addressed in the FM field are *interpretability* and *accuracy*. Of course, the ideal thing would be to satisfy both criteria to a high degree but, since they are contradictory issues, it is generally not possible. In this case, more priority is given to one of them (defined by the problem nature), leaving the other one in the background. Hence, two FM approaches arise depending on the main objective to be considered: LFM (interpretability) and PFM (accuracy).

Regardless of the approach, a common scheme is found in the existing literature to perform the FM:

1. Firstly, the main objective (interpretability or accuracy) is tackled defining a specific model structure to be used, thus setting the FM approach.
2. Then, the modeling components (model structure and/or modeling process) are improved by means of different mechanisms to define the desired ratio interpretability-accuracy.

This procedure results in four different possibilities (see Fig. 1): LFM with improved interpretability, LFM with improved accuracy, PFM with improved interpretability, and PFM with improved accuracy.

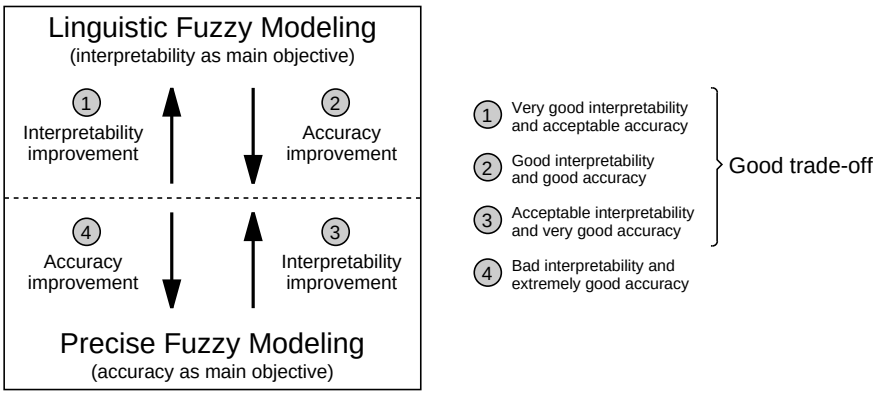


Fig. 1. Improvements of interpretability and accuracy in fuzzy modeling

Although historically more priority has been given to the accuracy, currently the search of a good balance between both criteria is increasing in importance. Indeed, a significative effort is being performed by several researchers proposing *improvement mechanisms* to compensate for the initial difference. Among the four said lines of work, clearly this philosophy is pursued by two of them: *LFM with improved accuracy* and *PFM with improved interpretability* (approaches 2 and 3 in Fig. 1, respectively).

Moreover, another interesting proposal is *LFM with improved interpretability* (approach 1 in Fig. 1). Although LFM uses a model structure with a high description power by itself, there are some problems (curse of dimensionality, excessive number of input variables or fuzzy rules, garbled fuzzy sets, etc.) that make it not to be as interpretable as desired and the need of interpretability improvements to restore the searched balance is justified.

Finally, the modus operandi of obtaining more accuracy in PFM (approach 4 in Fig. 1) does not pay attention to the comprehensibility of the model and acts close to black box techniques. This approach does not follow the original objective of FM and does not profit from the advantages that distinguish it from other modeling techniques. Although the approach is useful when only accuracy is required, it goes away from the aim of the present book.

This chapter is devoted to review different interpretability improvements that have been proposed to attain the desired balance. Thus, Sects. 4 and 5 show some mechanisms found in the recent literature to do so. In [27], an overview from a different point of view is explored by analyzing the in-

terpretability of several proposals instead of considering how the balance interpretability-accuracy is achieved.

3 Types of Fuzzy Rule-Based Systems

Before presenting the search of a balance interpretability-accuracy in FM, it seems that there is need to introduce the different kinds of FRBSs usually employed. It is a significant aspect to consider since depending on the rule structure used, an FRBS has itself a specific capability of description and approximation. The section is only focused on the FRBS types usually considered to improve their interpretability for the sake of a good trade-off.

3.1 Linguistic Fuzzy Rule-Based System

Also known as Mamdani-type FRBS [44,45], the linguistic FRBS constitutes the main tool to develop LFM. A crucial reason why this approach is worth considering is that it may remain verbally interpretable, playing the concept of linguistic variable [76] a central role. Linguistic FRBSs are formed by linguistic rules with the following structure:

IF X_1 is A_1 and ... and X_n is A_n
THEN Y_1 is B_1 and ... and Y_m is B_m ,

with X_i and Y_j being input and output linguistic variables respectively, and with A_i and B_j being linguistic labels with fuzzy sets associated defining their meaning. These linguistic labels will be taken from a global *semantic* defining the set of possible fuzzy sets used for each variable (Fig. 2 shows an example with triangular membership functions). This structure provides a natural framework to include expert knowledge in the form of fuzzy rules.

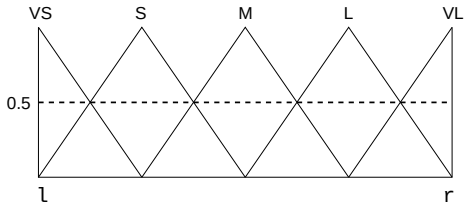


Fig. 2. Graphical representation of an example of the semantic considered for a variable, standing VS for *very small*, S for *small*, M for *medium*, L for *large*, and VL for *very large*, with $[l, r]$ being the corresponding variable domain

In these systems, the knowledge base (KB) – the component of the FRBS that stores the knowledge about the problem being solved – is composed of:

- the *rule base* (RB), constituted by the collection of linguistic rules themselves joined by means of the connective *also*, and
- the *data base* (DB), containing the term sets and the membership functions defining their semantics.

3.2 Takagi-Sugeno-Kang-type Fuzzy Rule-Based Systems

The system proposed by Takagi, Sugeno, and Kang [62,64] (shortly called TSK) differs from the linguistic one in the use of a different consequent structure. While linguistic rules consider a linguistic variable in the consequent, TSK-type fuzzy rules are based on representing the output variables as polynomial functions of the input variables, i.e.,

IF X_1 is A_1 and ... and X_n is A_n
THEN $Y_1 = p_1(X_1, \ldots, X_n)$ and ... and $Y_m = p_m(X_1, \ldots, X_n)$,

with $p_j(\cdot)$ being the polynomial function defined for the jth output variable. Using this fuzzy rule structure, the human interpretation on the action suggested by each rule is garbled but, on the contrary, the approximation capability is significantly increased. For this reason, TSK-type FRBSs are very useful in PFM.

3.3 Other Kinds of Fuzzy Rule-Based Systems

Other model structures may be considered apart from the two main types mentioned, some examples follow:

- The *singleton FRBS*, where the rule consequent takes a single real-valued number, may be considered as a particular case of the linguistic FRBS (the consequent is a fuzzy set where the membership function is one for a specific value and zero for the remaining ones) or of the TSK-type FRBS (the polynomial function of the consequent is a constant). Since the single consequent seems to be more easily interpretable than a polynomial function, the singleton FRBS may be used to develop LFM. Nevertheless, compared with the linguistic FRBS, the fact of having a different consequent value for each rule (no global semantic is used for the output variable) worsens the interpretability.
- The *fuzzy rule-based classification system*, which is an automatic classification system that uses fuzzy rules as knowledge representation tool. The classical fuzzy classification rule structure is the one that have a class label in the consequent part instead of the mentioned fuzzy set. Other alternative representations that consider a certainty degree for each rule or that include all the possible class labels with their corresponding certainty degrees in the consequent part are usually also considered.

- The *approximate FRBS*, which differs from the linguistic one in the direct use of fuzzy variable [1,7,15,36,61]. Each fuzzy rule thus presents its own semantic, i.e., the variables take different fuzzy sets as values and not linguistic terms from a global term set. The fuzzy rule structure is then as follows:

  IF X_1 is $\widehat{A}_1$ and ... and X_n is $\widehat{A}_n$
  THEN Y_1 is $\widehat{B}_1$ and ... and Y_m is $\widehat{B}_m$,

  with $\widehat{A}_i$ and $\widehat{B}_j$ being fuzzy sets. Since no global semantic is used in approximate FRBSs, these fuzzy sets can not be interpreted. This more flexible structure allows the model to be more accurate, being very appropriate to develop PFM.
  Other names have been proposed by different authors to designate approximate FRBSs. Among others, we may find FRBSs with *local fuzzy sets* [7], *rule-based* FRBSs [14], or *scatter-partitioning* FRBSs [22].

Moreover, different extensions to the FRBSs such as adding rule weights [20,74] or using disjunctive forms [13,42,24] have also been proposed.

4 Improving the Interpretability in Linguistic Fuzzy Modeling

A possibility to achieve a good trade-off between interpretability and accuracy is to perform an LFM process trying to obtain accurate initial models, and subsequently applying a process to improve the interpretability of the obtained model even at the expense of losing certain accuracy. To generate these accurate initial models in LFM, a large number of input variables, a great variety of linguistic terms, oversized RBs, or illegible fuzzy sets are usually considered. This section analyzes different mechanisms to improve the interpretability in these kinds of models. Moreover, there is always the chance of indirectly improving the accuracy in terms of generalization capability when removing the existing redundancies and inconsistencies.

4.1 Selecting Input Variables in the Model and/or in the Linguistic Rules

When managing high-dimensional problems with a large number of input variables, the RB suffers from an exponential growth in its size due to the homogeneous partitioning of the input and output spaces caused by the use of linguistic variables [4] and, therefore, a good interpretability is not guaranteed. Moreover, with an excessive number of input variables, every linguistic rule also loses part of its description ability since the understanding of the

condition to activate the rule comes more difficult. A solution to these disadvantages is to make an input variable selection process that reduces the number of variables used by the model.

Basically, we may distinguish between two variable selection processes:

- *Selecting input variables in the model* — This simplification task involves selecting a subset of input variables to be used in the model or, similarly, removing those input variables that do not significantly contribute to the FRBS performance.
  Usually, this variable (or feature) selection has been applied to LFM in classification [8,30,35,40,58,59], where a large number of variables is frequently tackled. Nevertheless, in [34] an interesting contribution is proposed for linguistic FRBSs including the variable selection within a more complex deriving process (with rule generation, DB tuning, and rule selection).
- *Selecting input variables in the linguistic rules* — Other innovative approach involves the selection of a subset of input variables for each rule. In this case, the antecedent length of each rule is variable avoiding the need of using all the variables involved in the system. It does not mean that a specific input variable ignored in a rule could not be used in another one. To manage with this structure in the inference process, a special linguistic term with a membership function with a value one in all the domain may be assigned to the ignored variables.
  In [35], this input variable selection at RB level is considered together with a global selection of the variables and a merging of the rules. On the other hand, the methods proposed in [10,26,43,68] obtain the most significative input variable for each rule during the learning process, instead of making an a priori variable selection.

It is important to emphasize that the fact of removing some input variables may cause the existence of several rules with identical antecedents (mainly when the selection is performed a posteriori) that could imply an inconsistence. In this case, the most usual solution is to merge these rules. This process is explained in the following section.

4.2 Selecting/Merging Linguistic Rules

Sometimes, an RB with an excessive size must be used to reach an acceptable accuracy degree in linguistic FRBSs. However, this effect is often caused by a deficient RB learning process (sometimes advisedly) with tendency to generate too many rules. Thus, in an RB we may find *redundant rules*, which do not contain relevant information and whose actions are covered by other rules; *erroneous rules*, which are wrong defined and distort the FRBS performance; and *conflictive rules*, which perturb the FRBS performance when coexist with others. Besides worsening the accuracy, an excessive number of rules makes difficult to understand the model behavior.

To face this problem, an *RB reduction process* can be developed by merging rules and/or selecting a subset of rules from a given RB to achieve the goal of minimizing the number of rules used while maintaining (or even improving) the FRBS performance. Indeed, depending on the criteria considered to reduce the RB, this process can be considered as a mechanism to improve not only the interpretability but also the accuracy.

The RB reduction is generally applied as a postprocessing stage, once an initial RB has been derived. We may distinguish between two approaches to reduce the fuzzy rule set size in order to obtain a *compact* RB:

- *Selecting linguistic rules* — It involves obtaining an optimized subset of rules from a previous RB by selecting some of them. We may find several methods to do so with different search algorithms in the specialized literature [15,16,23,29,32,33,38].
  In [39], an interesting heuristic rule selection procedure is proposed where, by means of statistical measures, a relevance factor is computed for each fuzzy rule composing the linguistic FRBSs to subsequently select the most relevant ones. The philosophy of ordering the rules with respect to an importance criterion and selecting a subset of them seems similar to the orthogonal transformation-methods used for TSK-type FRBSs [72] (explained in Sect. 5.1). Another heuristic rule selection procedure is proposed in [67].
- *Merging linguistic rules* — It is an alternative approach that reduces the RB by merging the existing linguistic rules. In [35], the authors propose to merge neighboring rules, i.e., linguistic rules where the linguistic terms used by the same variable in each rule are adjacent. The merge is performed in three different ways: using a new fuzzy set that groups the adjacent linguistic terms, merging the adjacent fuzzy sets if they are very similar, or giving the set of rules in disjunctive normal form. Another proposal is presented in [31], where a special consideration to the merging order is made.
  From a different point of view, the RB may be reduced by using a disjunctive form for the fuzzy rules that groups several rules within a more general expression, thus easing the interpretability. This approach is explained in next section.

4.3 Alternative Linguistic Rule Expressions

Another possibility to improve the interpretability of linguistic FRBSs is to use an alternative model structure to give a higher descriptive power to each rule. With this extended description we may represent a linguistic fuzzy model in a more compact structure with minor accuracy loss. To do that, the linguistic fuzzy rule expression is extended to make it more flexible. Some examples are shown in the following:

- *Disjunctive normal form (DNF)* — The DNF-type fuzzy rule has the following form [24]:

  IF X_1 is $\widetilde{A_1}$ and ... and X_n is $\widetilde{A_n}$ **THEN** Y is B

  where each input variable X_i takes as a value a set of linguistic terms $\widetilde{A_i} = \{A_{i1}\ or\ \ldots\ or\ A_{il_i}\}$, whose members are joined by a disjunctive operator, whilst the output variable remains a usual linguistic variable with a single label associated.

  This structure uses a more compact description that improves the interpretability. Moreover, the structure is a natural support to allow the absence of some input variables in each rule (simply making $\widetilde{A_i}$ be the whole set of linguistic terms). Several learning methods have been proposed following this rule structure [10,24–26,35,42,43,68].

- *Exception rules* — Another interesting possibility to represent a more interpretable and compact description is the use of exceptions [37]. In [6], the authors make a fine-tuning of the meaning of each linguistic rule by excluding a local region of its firing region. This consideration is especially useful when structures with multiple fuzzy input subspaces (e.g., the DNF one) are considered. An example follows:

  IF X_1 is {Big or Small} and X_2 is {Medium or Small} **THEN** Y is Big
  except **IF** X_1 is Small and X_2 is Medium

- *Union-rule configuration* — In [13], an attempt to palliate the *curse of dimensionality* problem (exponential growth in the number of rules when a large number of input variables are considered) is proposed by converting a multiple-input-variable linguistic rule into single-input-variable linguistic rules connected by the disjunction operator.

4.4 Linguistic Approximation

The linguistic approximation [19] lies in finding a linguistic description that represents a given fuzzy set. Given a linguistic FRBS where the DB has been automatically obtained or optimized, this procedure may be used in LFM to find an interpretation of the involved fuzzy sets to improve the comprehensibility of the model. During this linguistic approximation, a certain accuracy loss is assumed.

The linguistic approximation is usually performed with linguistic terms and sometimes linguistic modifiers are used as well. Some examples of methods that improve the interpretability of the model with linguistic approximation are [18,46,63].

5 Improving the Interpretability in Precise Fuzzy Modeling

The birth of more flexible FRBSs such as TSK or approximate ones also entails the eruption of PFM since the new structures allow the FM to achieve more accurate fuzzy models. This fact causes a shift from the seminal intent of FM and the modeling tasks with these kinds of FRBSs increasingly become black box processes.

Fortunately, nowadays there is a sense shared by several researches in the way of rescuing the good interpretability advantages offered by fuzzy systems. This interpretability consideration is usually attained by reducing the complexity of the model. On the other hand, approaches that improve the local description of the TSK-type fuzzy rules are also proposed. In the following subsections, some specific proposals are reviewed.

5.1 Ordering/Selecting TSK-type Fuzzy Rules

As mentioned in Sect. 4, an efficient way to improve the interpretability in FM is to select a subset of significative fuzzy rules that represent in a more compactly way the system to be modeled. Moreover, this selection of important rules has the interesting advantage of reducing the possible redundancy existing in the RB, thus improving the generalization capability of the system, i.e., its accuracy.

Recently, one of the most successful approaches to make such rule selection in TSK-type FRBS has been proposed by obtaining a subset of important fuzzy rules considering orthogonal transformations [47,48,55,57,66,69–72]. This mechanism is used to give an importance degree to each fuzzy rule, thus obtaining an ordering of them. Once they are sorted, the selection is achieved using only the most promising ones.

Given a previously defined RB, let us assume we have a matrix that allocates the firing degree of each rule for each training example considered:

$$F = \begin{bmatrix} f_{11} & f_{21} & \cdots & f_{r1} \\ f_{12} & f_{22} & \cdots & f_{r2} \\ \vdots & \vdots & \vdots & \vdots \\ f_{1N} & f_{2N} & \cdots & f_{rN} \end{bmatrix}$$

with f_{ij} being the normalized firing degree of the i-th rule when the j-th example is used, r the number of rules, and N the data set size. The relevance of each rule (column) may be analyzed by means of orthogonal transformations of this matrix.

The two orthogonal transformations and the most usual extensions to select a subset of TSK-type fuzzy rules are:

- *Orthogonal least-squares methods* (OLS) — The OLS-based method (whose first application to fuzzy rule selection was proposed in [66]) transforms the columns of the firing matrix F into a set of orthogonal basis vectors. With them, the individual error reduction ratio of each rule may be easily computed, thus defining its importance in the whole set of possible rules. Its main interest in system modeling is that it considers the output contribution of the rules to sort them. However, since the OLS-based method is guided by the approximation capabilities of the rules (fitting error) without paying attention to the premise structures, it is possible to give a high importance to redundant fuzzy rules with high firing degrees thanks to their contributions to the output [72].
  In [55], this drawback is faced considering the dependency between the current rule to be selected and the set of rules previously selected. If the firing vector corresponding to the current fuzzy rule is (or nearly is) a linear combination of the firing vectors corresponding to the previously selected rules, a low importance degree is assigned.
  In [47], another improvement to the basic OLS approach is made to consider the RB redundancy. Each time a new rule is selected, its similarity to the previously rules is analyzed. If it significantly differs from the others, the rule is added. Otherwise, the previously selected rule being similar to the current one is properly updated considering the latter.
- *Singular value decomposition and QR with column pivoting methods* (SVD-QR) — The first application of the SVD-QR method to the selection of the most important fuzzy rules was proposed in [48]. Firstly, the SVD algorithm obtains a factorization of the firing matrix F into a product of three matrices. The obtained information will determine the rules to be considered to construct the reduced RB. Then, QR with column pivoting is applied to determine the most important fuzzy rules.
  In [70], this method is improved disregarding the QR process to order the selected rules and obtaining this information directly from the SVD result (from the singular values). The main advantage of this improvement is its simplicity in terms of implementation and computational time consumption.
  In [55], the authors support the opinion that methods based on SVD fails to produce an importance ordering and they propose an optional solution only considering the QR process.

Generally, the final objective of orthogonal transformation-based methods is to order the candidate fuzzy rules for subsequently selecting the most important ones. Usually, the number of selected rules is established as a rule of thumb. However, in [57,71], statistical information criteria are used to automatically decide the number of rules to reduce the human intervention and consider a proper trade-off between simplicity of the model (interpretability) and data approximation (accuracy). Moreover, the use of orthogonal transformation has been used to perform other kinds of FM tasks with TSK-type FRBSs, such as the estimation of the consequent model parameters [47,73].

On the other hand, an interesting approach different from the orthogonal transformation one is proposed in [20] to make the TSK-type fuzzy rule selection. In this case, the model structure is extended incorporating an intensity parameter to each rule that allows the method to give different importance degrees during the inference process. These parameters could be considered as rule weights in linguistic FRBSs. This improvement makes the model more flexible giving it a higher approximation capability. To select the rules, the method considers an initial oversized RB and an iterative algorithm progressively removes the most redundant fuzzy rules.

Finally, we should say that an RB reduction is also indirectly attained when the fuzzy sets are merged or removed. The following section focuses on this approach.

5.2 Merging/Removing Fuzzy Sets in Precise Fuzzy Rule-Based Systems

Other successful way to obtain precise FRBSs (basically TSK or approximate ones) with a better interpretability is to reduce the number of rules by merging the fuzzy sets involved in the system. In this section, two different approaches are introduced in the following. While the former one tries to simplify TSK-type FRBSs, the latter one starts from a linguistic FRBSs with an excessive number of rules and fuzzy sets and generates a compact approximate FRBSs that is mostly equivalent.

The interpretability of TSK-type FRBSs may be improved by removing those fuzzy sets that, after an automatic adaptation and/or acquisition, do not significatively contribute to the model behavior. There are two effects caused by the fuzzy sets composing an FRBS that make the model unnecessarily more complex [52]:

- *Redundancy* — It refers to the coexistence of similar fuzzy sets representing compatible concepts. With these kinds of fuzzy sets, the model becomes more complex and difficult to be understandable (the distinguishability property [65] is not met).
- *Irrelevancy* — It is given when fuzzy sets with a constant membership degree equal to one, or close to it, are used. These kinds of fuzzy sets do not furnish relevant information.

To automatically detect these undesired fuzzy sets, the use of similarity measures between fuzzy sets has been proposed [51–53,56]. To properly use these measures for a RB reduction process, several properties – such as a similarity value of zero for nonoverlapping fuzzy sets, a value greater than zero for overlapping fuzzy sets, a value equal to one for equal fuzzy sets, and a measure independent of the scaling domain – must be satisfied [52]. For example, the following similarity measure between the fuzzy sets A and B for a discrete

domain is recommended to reduce the RB:

$$S(A,B) = \frac{\sum_{j=1}^{m} Min(\mu_A(x_j), \mu_B(x_j))}{\sum_{j=1}^{m} Max(\mu_A(x_j), \mu_B(x_j))}.$$

The process to simplify the model consists of two steps:

1. *Merging/removing fuzzy sets* — Those fuzzy sets with a high degree of similarity are merged to a unique fuzzy set that represent the collection of similar fuzzy sets. On the other hand, those irrelevant fuzzy sets – i.e., the ones with a high similarity degree to the universal set (a fuzzy set with a membership degree constantly one) – are removed. This step is called RB simplification [53].
2. *Merging fuzzy rules* — Moreover, it is interesting to mention that the fact of reducing the number of fuzzy sets in a variable fuzzy partition might result in rules with equal antecedents that can also be merged. This step is called RB reduction [53].

Hence, the precise fuzzy model go through an interpretability improvement (or complexity reduction) process that make it less complex (more compact) and more easily interpretable (more transparent).

Another interesting approach that also merges fuzzy sets is proposed in [61]. In this case, linguistic FRBSs with a bad interpretability are transformed into compact approximate FRBSs to develop PFM. Firstly, an iterative algorithm generates fuzzy partitions with a number of fuzzy sets large enough to achieve the desired accuracy degree, thus deliberately generating a linguistic RB with an excessive size. Subsequently, a merging process reduces it without losing accuracy by combining linguistic fuzzy rules that have adjacent fuzzy sets, thus obtaining a set of approximate fuzzy rules where each one has its own semantic.

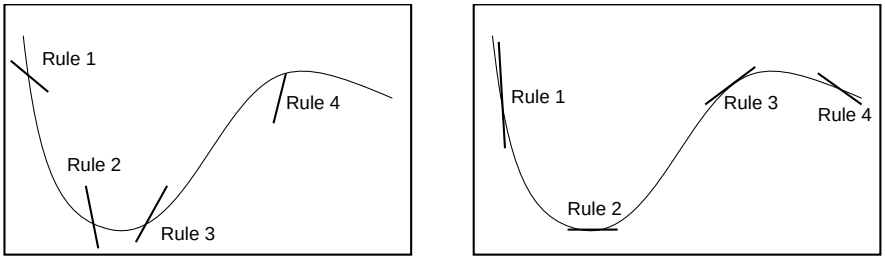
5.3 Exploiting the Local Description of TSK-type Fuzzy Rules

In system modeling, a TSK-type FRBS is usually considered as the combination of simple models (the rules) that describe local behaviors of the system to be modeled. Hence, insofar as each TSK-type fuzzy rule is either forced to have a smoother consequent polynomial function or to develop an isolated action, the interpretability will be improved. Several contributions follow these approaches to make TSK models more comprehensible:

- *Smoothing the consequent polynomial function* — For example, in [21] the author proposes a method that imposes several constraints to the weights involved in the polynomial function of each rule consequent:

$$w_0 = 0, \quad \sum_{j=1}^{n} w_j = 1, \quad w_j \geq 0 \ (j = 1, \ldots, n)$$

with w_j being the weight of the rule consequent polynomial function corresponding to the j-th input variable, w_0 the independent term, and n the number of input variables. Thanks to this, a convex combination of the input variables is performed, thus contributing to a better understanding of the model. Like in the previously mentioned contribution [20], a parameter associated with each rule is used to modulate its action.
Another approach that softens the consequent polynomial functions is proposed in [73]. To do that, an objective function that properly combines two criteria during the regression algorithm is used. On the one hand, the classical mean square error is considered as global measure error to evaluate the quality of the rule set, thus favoring the cooperation. On the other hand, a local error measure is used to induce competition among the rules. While the former criterion increases the accuracy, the latter allows the rules to describe each region better, thus improving the interpretability. Figure 3 shows two models with different description capabilities depending on the used polynomial functions.



(a) Model with garbled rules (b) Model with understandable rules

Fig. 3. The local interpretability of a TSK-type fuzzy model may be improved with smooth consequent polynomial functions

- *Isolating the fuzzy rule actions* — In [50], an study concludes that the description of each TSK-type fuzzy rule is improved when the overlapping between adjacent input fuzzy sets is reduced. This is because the performance region of a particular rule is more clearly defined by avoiding that other rules having a high firing degree in such an area. This approach is an alternative proposal to improve the local description of the TSK-type rules by designing the DB instead of the RB.

The proposal in [5] tries to englobe the two said philosophies for improving the local interpretability of TSK-type FRBSs. On the one hand, the use of a special type of membership function based on splines improves the local behavior of the fuzzy system by only firing the most immediate rules to the given input vector. On the other hand, the consequent polynomial structure is modified to interpret the coefficients as a Taylor series expansion around

the center of the corresponding rule (i.e., the vector containing the vertex of each membership function considered for each input variable in the fuzzy rule).

6 Concluding Remarks

The FM research developed in the last two decades was mainly focused on exploiting the flexibility of FM to obtain the maximum accuracy. During this evolution, the derivation methods were improved, the components to be designed were extended, and new model structures were proposed. This search of the accuracy usually set aside the interpretability of the obtained models.

However, we should remember the initial philosophy of fuzzy set theory directed to serve the bridge between the human understanding and the machine processing. In this challenge, the faculty of fuzzy models to express the behavior of the real system in a comprehensible manner acquires a great importance. This is why the current tendency in FM tries to find a better balance between interpretability and accuracy.

This equilibrium is attained from different perspectives. One of the things that attracts the eye is the fact that it is frequently performed by means of previous existing extensions, but used in a more rational and moderate way. Other times, however, new approaches explicitly proposed are considered. This chapter was aimed to present an introduction to the different tends recently proposed in the specialized literature to improve the interpretability degree of the fuzzy models with the objective of finding the desired trade-off.

The remaining 26 chapters contained in this volume are excellent works of research in the FM approach studied in this chapter and they properly represent the existing state-of-the-art.

References

1. R. Alcalá, J. Casillas, O. Cordón, and F. Herrera. Building fuzzy graphs: features and taxonomy of learning for non-grid-oriented fuzzy rule-based systems. To appear in *Journal of Intelligent and Fuzzy Systems*. Draft version available at `http://decsai.ugr.es/~casillas/`.
2. R. Babuška. *Fuzzy modeling for control*. Kluwer Academic, Norwell, MA, USA, 1998.
3. R. Babuška, H. Bersini, D.A. Linkens, D. Nauck, G. Tselentis, and O. Wolkenhauer. Future prospects for fuzzy systems and technology. ERUDIT Newsletter Vol. 6, No. 1. Aachen, Germany, 2000. Available at `http://www.erudit.de/erudit/newsletters/news_61/page5.htm`.
4. 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.

5. M. Bikdash. A highly interpretable form of Sugeno inference systems. *IEEE Transactions on Fuzzy Systems*, 7(6):686–696, 1999.
6. P. Carmona, J.L. Castro, and J.M. Zurita. Learning maximal structure fuzzy rules with exceptions. In *Proceedings of the 2nd International Conference in Fuzzy Logic and Technology*, pages 113–117, Leicester, UK, 2001.
7. B. Carse, T.C. Fogarty, and A. Munro. Evolving fuzzy rule based controllers using genetic algorithms. *Fuzzy Sets and Systems*, 80:273–294, 1996.
8. J. Casillas, O. Cordón, M.J. del Jesus, and F. Herrera. Genetic feature selection in a fuzzy rule-based classification system learning process for high dimensional problems. *Information Sciences*, 136(1-4):169–191, 2001.
9. J. Casillas, O. Cordón, and F. Herrera. Can linguistic modeling be as accurate as fuzzy modeling without losing its description to a high degree? Technical Report #DECSAI-00-01-20, Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain, 2000. Available at `http://decsai.ugr.es/~casillas/`.
10. J.L. Castro, J.J. Castro-Schez, and J.M. Zurita. Learning maximal structure rules in fuzzy logic for knowledge acquisition in expert systems. *Fuzzy Sets and Systems*, 101(3):331–342, 1999.
11. J.L. Castro, C.J. Mantas, and J.M. Benítez. Interpretation of artificial neural networks by means of fuzzy rules. *IEEE Transactions on Neural Networks*, 13(1):101–116, 2002.
12. Z. Chi, H. Yan, and T. Pham. *Fuzzy algorithms with application to image processing and pattern recognition*. World Scientific, Singapore, 1996.
13. W.E. Combs and J.E. Andrews. Combinatorial rule explosion eliminated by a fuzzy rule configuration. *IEEE Transactions on Fuzzy Systems*, 6(1):1–11, 1998.
14. M.G. Cooper and J.J. Vidal. Genetic design of fuzzy controllers: the cart and jointed pole problem. In *Proceedings of the 3rd IEEE International Conference on Fuzzy Systems*, pages 1332–1337, Piscataway, NJ, USA, 1994.
15. O. Cordón and F. Herrera. A three-stage evolutionary process for learning descriptive and approximate fuzzy logic controller knowledge bases from examples. *International Journal of Approximate Reasoning*, 17(4):369–407, 1997.
16. O. Cordón and F. Herrera. A proposal for improving the accuracy of linguistic modeling. *IEEE Transactions on Fuzzy Systems*, 8(3):335–344, 2000.
17. D. Driankov, H. Hellendoorn, and M. Reinfrank. *An introduction to fuzzy control*. Springer-Verlag, Heidelberg, Germany, 1993.
18. A. Dvořák. On linguistic approximation in the frame of fuzzy logic deduction. *Soft Computing*, 3(2):111–116, 1999.
19. F. Eshragh and E.H. Mamdani. A general approach to linguistic approximation. In E.H. Mamdani and B.R. Gaines, editors, *Fuzzy Reasoning and its Applications*, pages 169–187. Academic Press, London, UK, 1981.
20. A. Fiordaliso. Autostructuration of fuzzy systems by rules sensitivity analysis. *Fuzzy Sets and Systems*, 118(2):281–296, 2001.
21. A. Fiordaliso. A constrained Takagi-Sugeno fuzzy system that allows for better interpretation and analysis. *Fuzzy Sets and Systems*, 118(2):307–318, 2001.
22. B. Fritzke. Incremental neuro-fuzzy systems. In B. Bosacchi, J.C. Bezdek, and D.B. Fogel, editors, *Proceedings of the International Society for Optical Engineering: Applications of Soft Computing*, volume 3165, pages 86–97, 1997.
23. A.F. Gómez-Skarmeta and F. Jiménez. Fuzzy modeling with hybrid systems. *Fuzzy Sets and Systems*, 104(2):199–208, 1999.

24. A. González and R. Pérez. Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems*, 96(1):37–51, 1998.
25. A. González and R. Pérez. SLAVE: a genetic learning system based on an iterative approach. *IEEE Transactions on Fuzzy Systems*, 7(2):176–191, 1999.
26. A. González and R. Pérez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 31(3):417–425, 2001.
27. S. Guillaume. Designing fuzzy inference systems from data: an interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9(3):426–443, 2001.
28. K.M. Hangos, editor. *Special issue on grey box modelling*, volume 9(6) of *International Journal of Adaptive Control and Signal Processing*. John Wiley & Sons, New York, NY, USA, 1995.
29. F. Herrera, M. Lozano, and J.L. Verdegay. A learning process for fuzzy control rules using genetic algorithms. *Fuzzy Sets and Systems*, 100:143–158, 1998.
30. T.-P. Hong and J.-B. Chen. Finding relevant attributes and membership functions. *Fuzzy Sets and Systems*, 103(3):389–404, 1999.
31. T.-P. Hong and C.-Y. Lee. Effect of merging order on performance of fuzzy induction. *Intelligent Data Analysis*, 3(2):139–151, 1999.
32. H. Ishibuchi, T. Murata, and I.B. Türkşen. Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems. *Fuzzy Sets and Systems*, 89(2):135–150, 1997.
33. H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Selecting fuzzy if-then rules for classification problems using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(3):260–270, 1995.
34. Y. Jin. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement. *IEEE Transactions on Fuzzy Systems*, 8(2):212–221, 2000.
35. A. Klose, A. Nurnberger, and D. Nauck. Some approaches to improve the interpretability of neuro-fuzzy classifiers. In *Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing*, pages 629–633, Aachen, Germany, 1998.
36. L. Koczy. Fuzzy if ... then rule models and their transformation one another. *IEEE Transactions on Systems, Man, and Cybernetics*, 26(5):621–637, 1996.
37. A. Krone and H. Kiendl. Automatic generation of positive and negative rules for two-way fuzzy controllers. In *Proceedings of the Second European Congress on Intelligent Techniques and Soft Computing*, volume 1, pages 438–447, Aachen, Germany, 1994. Verlag Mainz.
38. A. Krone, P. Krause, and T. Slawinski. A new rule reduction method for finding interpretable and small rule bases in high dimensional search spaces. In *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*, pages 693–699, San Antonio, TX, USA, 2000.
39. A. Krone and H. Taeger. Data-based fuzzy rule test for fuzzy modelling. *Fuzzy Sets and Systems*, 123(3):343–358, 2001.
40. H.-M. Lee, C.-M. Chen, J.-M. Chen, and Y.-L. Jou. An efficient fuzzy classifier with feature selection based on fuzzy entropy. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 31(3):426–432, 2001.
41. P. Lindskog. Fuzzy identification from a grey box modeling point of view. In H. Hellendoorn and D. Driankov, editors, *Fuzzy model identification*, pages 3–50. Springer-Verlag, Heidelberg, Germany, 1997.

42. L. Magdalena. Adapting the gain of an FLC with genetic algorithms. *International Journal of Approximate Reasoning*, 17(4):327–349, 1997.
43. L. Magdalena and F. Monasterio-Huelin. A fuzzy logic controller with learning through the evolution of its knowledge base. *International Journal of Approximate Reasoning*, 16(3):335–358, 1997.
44. E.H. Mamdani. Applications of fuzzy algorithms for control a simple dynamic plant. *Proceedings of the IEE 121*, 12:1585–1588, 1974.
45. E.H. Mamdani and S. Assilian. An experiment in linguistic systhesis with fuzzy logic controller. *International Journal of Man-Machine Studies*, 7:1–13, 1975.
46. J.G. Marín-Blázquez, Q. Shen, and A.F. Gómez-Skarmeta. From approximative to descriptive models. In *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*, pages 829–834, San Antonio, TX, USA, 2000.
47. P.A. Mastorocostas, J.B. Theocharis, and V.S. Petridis. A constrained orthogonal least-squares method for generating TSK fuzzy models: application to short-term load forecasting. *Fuzzy Sets and Systems*, 118(2):215–233, 2001.
48. G.C. Mouzouris and J.M. Mendel. A singular-value-QR decomposition based method for training fuzzy logic systems in uncertain environments. *Journal of Intelligent and Fuzzy Systems*, 5:367–374, 1997.
49. W. Pedrycz, editor. *Fuzzy modelling: paradigms and practice*. Kluwer Academic, Norwell, MA, USA, 1996.
50. A. Riid and E. Rüstern. Interpretability versus adaptability in fuzzy systems. *Proceedings of the Estonian Academy of Sciences. Engineering*, 49(2):76–95, 2000.
51. H. Roubos and M. Setnes. Compact and transparent fuzzy models and classifiers through iterative complexity reduction. *IEEE Transactions on Fuzzy Systems*, 9(4):516–524, 2001.
52. M. Setnes, R. Babuška, U. Kaymak, and H.R. van Nauta Lemke. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 28(3):376–386, 1998.
53. M. Setnes, R. Babuška, and H.B. Verbruggen. Complexity reduction in fuzzy modeling. *Mathematics and Computers in Simulation*, 46(5-6):509–518, 1998.
54. M. Setnes, R. Babuška, and H.B. Verbruggen. Rule-based modeling: precision and transparency. *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, 28(1):165–169, 1998.
55. M. Setnes and H. Hellendoorn. Orthogonal transforms for ordering and reduction of fuzzy rules. In *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*, pages 700–705, San Antonio, TX, USA, 2000.
56. M. Setnes and H. Roubos. GA-fuzzy modeling and classification: complexity and performance. *IEEE Transactions on Fuzzy Systems*, 8(5):509–522, 2000.
57. L.I.U. Shi-Rong and Y.U. Jin-Shou. Model construction optimization for a class of fuzzy models. *Chinese Journal of Computers*, 24(2):164–172, 2001.
58. R. Silipo and M. Berthold. Discriminative power of input features in a fuzzy model. In D. Hand, J. Kok, and M. Berthold, editors, *Advances in Intelligent Data Analysis (IDA–99)*, volume LNCS 1642, pages 87–98. Springer-Verlag, Heidelberg, Germany, 1999.
59. R. Silipo and M. Berthold. Input features' impact on fuzzy decision processes. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 30(6):821–834, 2000.
60. T. Söderström and P. Stoica. *System identification*. Prentice Hall, Engleweed Cliffs, NJ, USA, 1989.

61. T. Sudkamp, J. Knapp, and A. Knapp. Refine and merge: generating small bases from training data. In *Proceedings of the 9th IFSA World Congress and the 20th NAFIPS International Conference*, pages 197–202, Vancouver, Canada, 2001.
62. M. Sugeno and G.T. Kang. Structure identification of fuzzy model. *Fuzzy Sets and Systems*, 28:15–33, 1988.
63. M. Sugeno and T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1):7–31, 1993.
64. T. Takagi and M. Sugeno. Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15:116–132, 1985.
65. J. Valente de Oliveira. Semantic constraints for membership function optimization. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 29(1):128–138, 1999.
66. L.-X. Wang and J.M. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least squares learning. *IEEE Transactions on Neural Networks*, 3:807–814, 1992.
67. X. Wang and J. Hong. Learning optimization in simplifying fuzzy rules. *Fuzzy Sets and Systems*, 106(3):349–356, 1999.
68. N. Xiong and L. Litz. Fuzzy modeling based on premise optimization. In *Proceedings of the 9th IEEE International Conference on Fuzzy Systems*, pages 859–864, San Antonio, TX, USA, 2000.
69. Y. Yam, P. Baranyi, and C.-T. Yang. Reduction of fuzzy rule base via singular value decomposition. *IEEE Transactions on Fuzzy Systems*, 7(2):120–132, 1999.
70. J. Yen and L. Wang. An SVD-based fuzzy model reduction strategy. In *Proceedings of the 5th IEEE International Conference on Fuzzy Systems*, pages 835–841, New Orleans, LA, USA, 1996.
71. J. Yen and L. Wang. Application of statistical information criteria for optimal fuzzy model construction. *IEEE Transactions on Fuzzy Systems*, 6(3):362–372, 1998.
72. J. Yen and L. Wang. Simplifying fuzzy rule-based models using orthogonal transformation methods. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 29(1):13–24, 1999.
73. J. Yen, L. Wang, and C.W. Gillespie. Improving the interpretability of TSK fuzzy models by combining global learning and local learning. *IEEE Transactions on Fuzzy Systems*, 6(4):530–537, 1998.
74. W. Yu and Z. Bien. Design of fuzzy logic controller with inconsistent rule base. *Journal of Intelligent and Fuzzy Systems*, 2:147–159, 1994.
75. L.A. Zadeh. Outline of a new approach to the analysis of complex systems and desision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, 3:28–44, 1973.
76. L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. Parts I, II and III. *Information Science*, 8, 8, 9:199–249, 301–357, 43–80, 1975.