Open-loop fuzzy modeling: a comparison and a proposal

P. Carmona
Depto. Informática
Univ. de Extremadura
06071 Badajoz, Spain
pablo@unex.es

J.L. Castro
Depto. C.C. e I.A.
Univ. de Granada
18071 Granada, Spain
castro@decsai.ugr.es

J.M. Zurita
Depto. C.C. e I.A.
Univ. de Granada
18071 Granada, Spain
zurita@decsai.ugr.es

Abstract

In open-loop fuzzy modeling (OLFM), the examples are translated into a set of rules in only one pass. We analyse the behaviors and features of different methods for OLFM combining two strategies for identifying fuzzy rules from certainty degrees with two well-known measures of that degrees. The methods are analysed experimentally by considering the model error when applied to several systems. A new method combining the best components is proposed and the results are analysed, showing its improvements in accuracy and computational cost.

Keywords: Fuzzy modeling, rule certainty degrees, experimental analysis.

1 Introduction

Fuzzy learning methods usually identify the rule base by selecting the most promising rules from the training examples in only one pass [1, 3, 4]. This rule base will constitute either the definitive rule base or, in more complex techniques, a preliminary rule base that is refined later. That one–pass modeling process —which we will call OLFM— partially relies its success on the strategy followed to obtain the rule base, together with the certainty measure considered.

Two strategies and two well-known and widely-used certainty degrees measures have been considered, and a comparison of the four possible methods has been performed. In the analysis, the fuzzy model was dissected, obtaining different suitabilities for different methods depending on the group of rules being identified. This outcome leads us to propose a Mixed Method keeping the best behaviors which obtains a double gain: accuracy and computational cost.

The next two sections present the strategies and the certainty degree measures to be compared, respectively. Then, Section 4 expounds the results obtained in the comparison of the methods. Finally, Section 5 describes the MM proposed by the authors and its identification results are compared with the previous ones.

2 Strategies for OLFM

We consider a set of MISO rules:

$$R^{i_1 \ldots i_n}_{LY_i} : LX_{1,i_1}, \ldots, LX_{n,i_n} \to LY_i$$

where $LX_{k,i_k} \in \widetilde{\mathcal{X}_k}$ and $LY_i \in \widetilde{\mathcal{Y}}$, and a set of examples $E = \{e^1, \ldots, e^m\}$, $e^i = ([x^i_1, ..., x^i_n], y^i)$.

The first strategy is an input space oriented strategy (ISS) that tries to identify the best rule by examining each input fuzzy region. It is described by the following algorithm:

For each fuzzy region $\mathbf{LX} = [LX_{1,i_1}, \ldots, LX_{n,i_n}]$:
  Set $\omega max = 0$.
  For each output fuzzy label LY_i:
    Obtain the rule certainty degree ω_i from E.
    If $\omega_i > \omega max$ then
      set $\omega max = \omega_i$ and $max = i$.
  If $\omega max > 0$, then assign LY_{max} to $\mathbf{LX}$.

A certainty degree higher than zero means some confidence that the rule is correct based on E. If such a rule does not exist, no one will be assigned to the corresponding input fuzzy region.

The second strategy is an example oriented strategy (ES), which selects the best translation from examples into rules. It is described as follows:

Set $\omega max^{i_1 \dots i_n} = 0$, for all fuzzy input regions.
For each example e^j in E:
  Set up the *main rule* of e^j as $R_{LY_i}^{i_1 \dots i_n}$, where $\mu_{LX_{k,i_k}}(x_k^j) = \max_l \mu_{LX_{k,l}}(x_k^j)$, for all k, and $\mu_{LY_i}(y^j) = \max_l \mu_{LY_l}(y^j)$.
  Obtain its certainty degree ω_i from E.
  If $\omega_i > \omega max^{i_1 \dots i_n}$, then set $\omega max^{i_1 \dots i_n} = \omega_i$ and assign LY_i to $[LX_{1,i_1}, \dots, LX_{n,i_n}]$.

3 Certainty measures

The analysed measures were the Wang and Mendel's measure [5] and the Ishibuchi's measure [2] extended for dealing with fuzzy consequents.

Definition 1 (WMM) *Since an ES is used in [5], each rule has an example e^j associated and WMM is defined as:*

$$\omega_{WM}(R_{LY_i}^{i_1 \dots i_n}) = \mu_{LX_{1,i_1}}(x_1^j) \times \cdots \cdots \times \mu_{LX_{n,i_n}}(x_n^j) \times \mu_{LY_i}(y^j) \quad (1)$$

Since ISS does not associate any specific example to each rule, following the idea of taking the best covered rule in each region, WMM is defined as:

$$\omega_{WM}(R_{LY_i}^{i_1 \dots i_n}) = \max_{e^j \in E} \{\mu_{LX_{1,i_1}}(x_1^j) \times \cdots \cdots \times \mu_{LX_{n,i_n}}(x_n^j) \times \mu_{LY_i}(y^j)\} \quad (2)$$

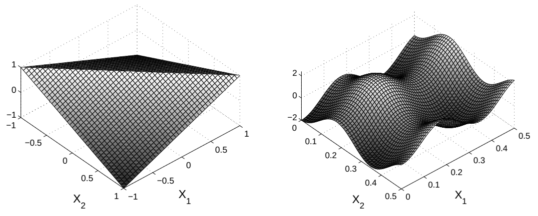
Definition 2 (IM) *IM is defined as:*

$$\omega_I(R_{LY_i}^{i_1 \dots i_n}) = \frac{\beta(R_{LY_i}^{i_1 \dots i_n}) - \overline{\beta}(R_{LY_i}^{i_1 \dots i_n})}{\sum_{k=1}^{q} \beta(R_{LY_k}^{i_1 \dots i_n})} \quad (3)$$

$$\beta(R_{LY_i}^{i_1 \dots i_n}) = \sum_{e_j \in E} \mu_{LX_{1,i_1}}(x_1^j) \times \cdots \cdots \times \mu_{LX_{n,i_n}}(x_n^j) \times \mu_{LY_i}(y^j) \ , \quad (4)$$

$$\overline{\beta}(R_{LY_i}^{i_1 \dots i_n}) = \sum_{\substack{k=1 \\ (k \neq i)}}^{q} \frac{\beta(R_{LY_k}^{i_1 \dots i_n})}{q-1} \ , \quad (5)$$

where q is the number of labels in $\widetilde{\mathcal{Y}}$.



(a) f_1 (b) f_2

Figure 1: Output surfaces of functions f_1 and f_2.

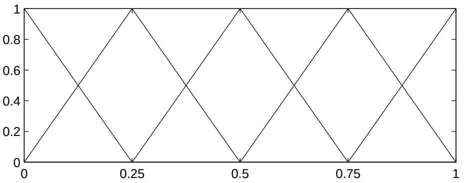


Figure 2: Fuzzy domains.

The main difference between both measures is the number of examples involved in the calculus: WMM takes into account only the most representative example, whereas IM involves all the examples covering the rule antecedent.

4 Comparison

In order to compare the four resultant methods, several experiments were performed over two functions with different complexity:

$$f_1(x_1, x_2) = 1 - |x_1 - x_2|, [-1, 1]^2 \rightarrow [-1, 1] \quad (6)$$

$$f_2(x_1, x_2) = x_1 + x_2 - cos(18x_1) - cos(18x_2), [0, 0.5]^2 \rightarrow [-2, 2.4] \quad (7)$$

whose surfaces are shown in Figure 1. All fuzzy domains consist of 5 equally spaced triangular membership functions as shown in Figure 2 once they are normalized.

The experiment consisted in the identification of the systems with randomly generated training sets varying from 1 to 1000 examples in increasing steps as the size increases. For each training set size, every method was run 100 times with a different training set each.

Respecting the total number of identified rules, ISS always identifies at least the same number of rules than ES, since the latter assigns a rule to an input region only if that rule is a main rule for

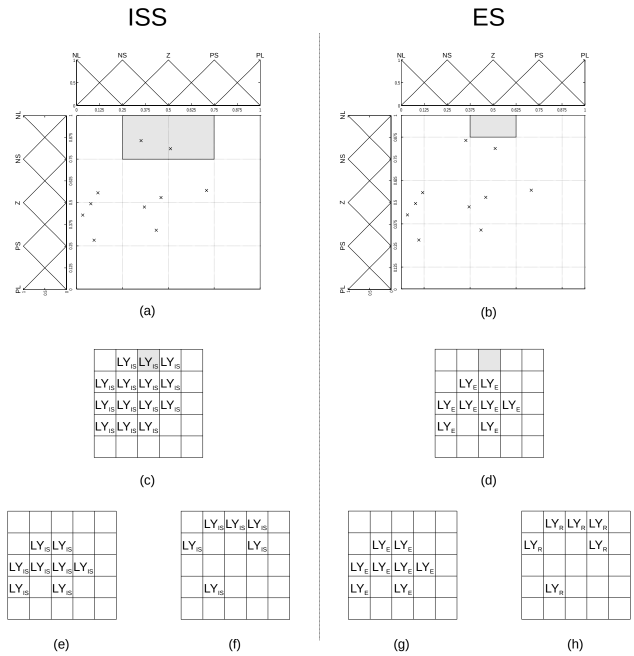


Figure 3: Differences between ISS and ES.

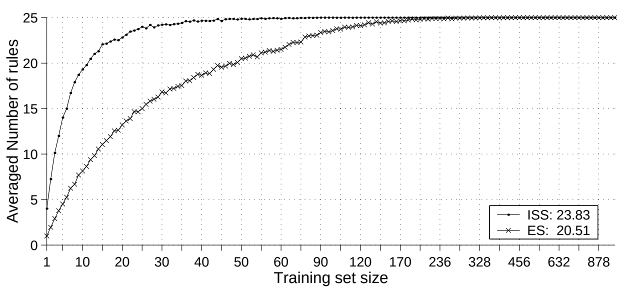


Figure 4: Number of rules identified with ISS and ES.

some example, whereas in the former a rule will be identified if some example exists matching its antecedent in any degree. Figures 3(a-d) illustrate this effect: for example, the rule in $[Z, NL]$ (dark zone) is identified with ISS since two examples cover that region, whereas it is not with ES since the region is not a *main region* for none of both.

Figure 4 presents the experimental results,[1] where ISS achieved a complete rule base with nearly 70 examples, whereas ES needed about 400 examples for it. However, as it will be shown below, it does not mean that ISS achieves, in general, a better model than ES.

[1] The results are the sames for each function and measure considered, since they only depend on the distribution of the examples in the input space.

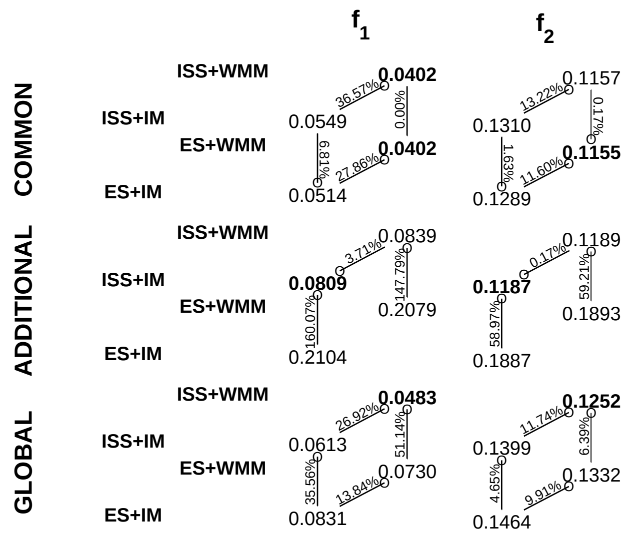


Figure 5: Model errors with the methods.

Regarding this, it would be interesting to analyse if the higher number of rules identified with ISS justifies itself better accuracy results, even if these additional rules were selected randomly. With this aim, two groups of rules were differentiated:

- *Rules in common regions (CR)*, located in regions where both strategies identify rules.
- *Rules in additional regions (AR)*, located in regions where only ISS identifies rules.

Since AR is empty for ES, rules with randomly selected consequents were considered for this case, thus allowing to determine if ISS identifies rules in AR better than using a random selection.

Figures 3(c-h) show the dissection performed to evaluate each group: for CR, two rule bases are set up which contain the rules identified by ISS and ES in CR (Figures 3e and 3g); for AR, a rule base contains the rules identified with ISS in AR (Figure 3f) and another the rules in the same location with random consequents (Figure 3h).

The model errors are summarized in Figure 5 — the mean absolute error over all the training set sizes and runs—, along with the superiority relation between measures and strategies (%).

Focusing on CR, when applying IM, the results shows that ES provides higher accuracy than ISS, which amounts that considering more rules than the main rule in a CR distorts, in general, the appropriate selection. In this respect, it must be noticed that a CR contains examples in its central zone and, therefore, it is expected that these examples are representative of the model output in that region. Thus, ISS, which also considers rules without central examples, along with IM, which selects the rule with a high *averaged* covering, can lead to select rules that are not represented by the examples located in the central zone if most of them are in the periphery of the region.

When applying WMM, the differences between ISS and ES diminish. Now, the certainty degree of each non-main rule in ISS equals to its *best* covering degree, which likely is lower than those of main rules, which are covered at least by an example in a high degree: the one who yields it.

Finally, WMM provides much better accuracy than IM in CR. It is basically due to the same reason that leads ES to outperform ISS: considering only the information provided by the "best" example. The attention that IM pays to low covering examples when dealing with CR leads to disturbances in the selection of the correct rule.

Respecting the AR, ISS reports better results than a mere random selection.[2] Therefore, ISS must be used when facing regions in which only non-main rules exist. When comparing measures, now IM outperforms WMM.[3] The reason lies in that the "best-covering" examples in AR are now not so good examples, and then, extending the evaluation to all the covering examples attenuate the negative effect of that best-covering example.

Regarding the global behaviors, ISS+WMM provides the best accuracy, due to the clear superiority of WMM over IM in CR and of ISS over ES in AR. However, two concealed results highlighted during the analysis are the general outperformance of ES in CR and of IM in AR.

In terms of computational cost, ISS is expensive, specially with high dimensional systems, since it grows exponentially with the number of variables.

[2]Since error of ES in AR is independent of the system complexity —which is not the case for ISS— the superiority of ISS over ES is higher in f_1 than in f_2.

[3]In this case, only ISS is analysed, since in ES the rules are generated randomly.

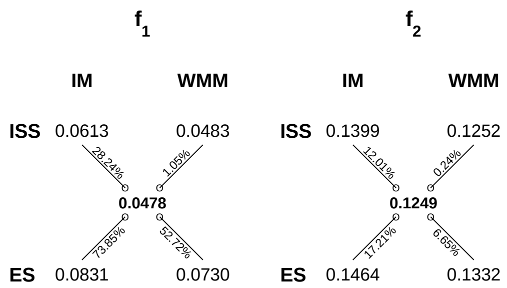


Figure 6: Model errors with MM.

It does not happen with ES, which mainly depends on the training set size. Between ES methods, ES+IM is specially sensible to this size, since IM considers all the covering examples each time an example is translated into a main rule.

5 Mixed Method: a proposal

The results obtained in previous section lead us to propose a Mixed Method groping the different best performances. It is described as follows:

Perform the identification using ES+WMM.
For each fuzzy region **LX** not identified yet:
    Set $\omega max = 0$.
    For each output fuzzy label LY_i:
        Obtain $\omega_I(R^{i_1 \ldots i_n}_{LY_i})$
        If $\omega_i > \omega max$ then
            set $\omega max = \omega_i$ and $max = i$.
    If $\omega max > 0$, assign LY_{max} to **LX**.

When applying the experiments to MM, the errors in CR and AR equal, obviously, to the ones of ES+WMM in CR and of ISS+IM in AR. Regarding the global errors, Figure 6 summarizes their results (centered) contrasted with the other methods, showing that MM outperforms all of them. The detailed results are shown in Figure 7, where MM shows similar error than ISS+WMM in spite of being a mixture of ES+WMM and ISS+IM.[4]

Moreover, although MM begins with a cost similar to ISS, it decreases as the training set size increases, approximating to a pure ES+WMM since AR disappear (Figure 8). Thus, MM provides an accuracy/cost trade off: with few examples, its higher computational cost respecting ES+WMM is compensated with its higher accuracy; with

[4]Similar results are obtained for both functions.

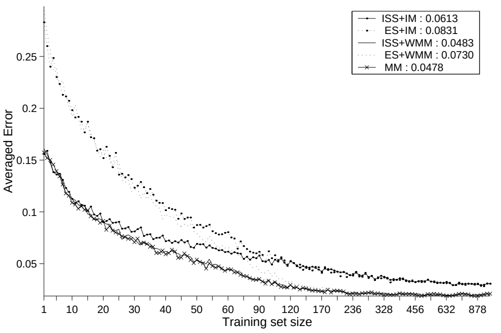


Figure 7: Detailed model errors.

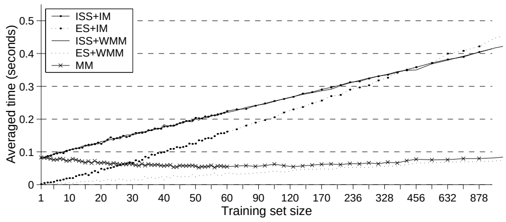


Figure 8: Computational cost.

more examples, the accuracy is similar (slightly better) than ISS+IM but with a much lower cost.

6 Conclusions

The main result lies in the fact that the differentiation of several groups of rules reveals good performances hidden from a global perspective. In general, ES performs better in regions where *main* rules exist, and ISS, which identifies regions not identified by ES, performs better in that regions than a random selection. Besides, WMM provides better results in CR and IM do it in AR. Thus, MM is proposed, in order to grasp the benefits of each strategy and measure. The combination of ES and WMM in CR and ISS and IM in AR provides the best accuracy and computational results. It can be helpful to researchers using any form of OLFM in their works.

References

[1] O. Cordon, F. Herrera, and P. Villar. Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Trans. Fuzzy Syst.*, 9(4):667–674, 2001.

[2] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka. Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Sets Syst.*, 65:237–253, 1994.

[3] J.F. Martins. New trends in recognizing experimental drives: Fuzzy logic and formal language theories. *IEEE Trans. Fuzzy Syst.*, 9(1):68–87, 2001.

[4] I. Rojas, H. Pomares, J. Ortega, and A. Prieto. Self-organized fuzzy system generation from training examples. *IEEE Trans. Fuzzy Syst.*, 8(1):23–36, 2000.

[5] L.X. Wang and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst., Man, Cybern.*, 22(6):1415–1427, 1992.