On Interpretability of Fuzzy Models Based on Conciseness Measure

T. Furuhashi
Dept. of Information Enegineering
Mie University
1515 Kamihama-cho, Tsu 514-8507, Japan
Tel.& Fax. +81-59-231-9456, E-mail: furuhashi@pa.info.mie-u.ac.jp

T. Suzuki
Dept. of Information Electronics, Nagoya University, Japan

Abstract

Fuzzy modeling is a method to describe input-output relationships of unknown systems using fuzzy inference. Interpretability is one of the indispensable features of fuzzy models. This paper discusses the interpretability of fuzzy model with/without prior knowledge about the target system. Without prior knowledge, conciseness of fuzzy model helps humans to interpret its input-output relationships. In the case where a human has the knowledge in advance, an interpretable model could be the one that explicitly explains his/her knowledge. This paper defines the conciseness of fuzzy models, and formulates the conciseness measure. Experimental results show that the obtained concise model has the essential interpretable feature. The results also show that human's knowledge changes the most interpretable model from the most concise model.

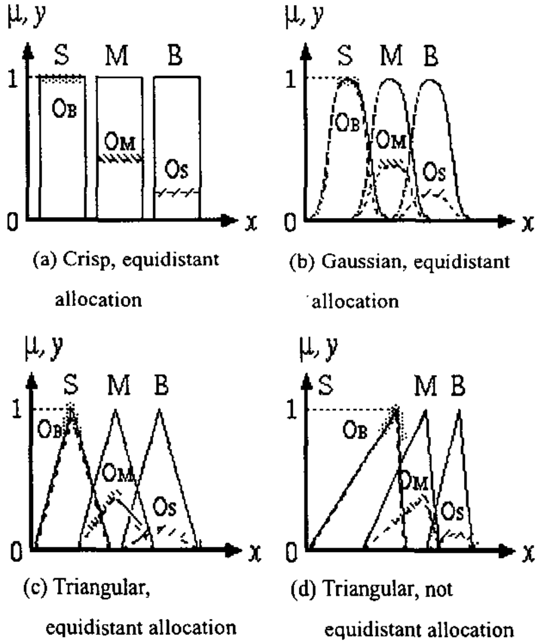
I. INTRODUCTION

Problems of describing input-output relationships of unknown systems from data have attracted much attention in many fields. Fuzzy modeling is one of the effective tools for solving the problems. The distinguishing feature of fuzzy model is in that it is interpretable. However, there have been few reports on quantitative analysis of the interpretability of fuzzy models.

The interpretability of fuzzy models has been evaluated simply by the number of fuzzy rules, the number of membership functions [1], or the degree of freedom term of AIC (Akaike's Information Criterion) [2, 3]. Matsushita, Furuhashi, et al. [1] discussed hierarchical fuzzy modeling for identifying interpretable fuzzy models. However, automatically derived fuzzy models are not often linguistically interpretable, as recognized in [4 - 6]. The interpretability of fuzzy models also depends on other factors such as shape/allocation of membership functions, and more on interpreter's prior knowledge.

This paper studies the interpretability of fuzzy models by separating the cases where prior knowledge is available or not. This paper defines conciseness of fuzzy models that is an essential factor for the interpretability. Without prior knowledge, input-output relationships of a "concise" fuzzy model are easy for a human to understand. This paper introduces De Luca and Termini's fuzzy entropy [7] as a conciseness measure that evaluates the shape of membership function. This paper also presents a new measure derived on the analogy of relative entropy. This new measure is also a conciseness measure that evaluates the deviation of allocation of membership functions on the universe of discourse. A combination of these two measures is shown to be a good conciseness measure.

In the case where a human has *a priori* knowledge about the target system in advance, a concise model is not always interpretable. In this case, an interpretable model could be the one that explicitly explains human's knowledge. Experimental results show that the obtained concise model has the essential interpretable feature. The results also show that human's knowledge changes the most interpretable model from the most concise model.



(a) Crisp, equidistant allocation

(b) Gaussian, equidistant allocation

(c) Triangular, equidistant allocation

(d) Triangular, not equidistant allocation

Fig.1 Membership functions and outputs as granules

0-7803-7293-X/01/$17.00 © 2001 IEEE

II. Interpretability and Conciseness

Interpretability of fuzzy models heavily depends on human's prior knowledge. If we have profound knowledge about the target system, an interpretable model could be the one that makes our knowledge explicit. Even though the fuzzy model had many parameters and the input-output relationships were highly non-linear, our knowledge helps us to interpret the relationships. Even a concise model is not interpretable, if it does not fit into our prior knowledge.

In the case where we have no *a priori* knowledge, a concise fuzzy model could be easy for a human to interpret. Assuming that four types of fuzzy models are given as shown in Fig.1. All the models in this figure are single-input single-output ones. Fig.1(a) shows a case where crisp membership functions S, M, and B are allocated equidistantly on the universe of discourse in the antecedent. The output is depicted with granules O_B, O_M, and O_S. This model is interpreted as

If x is S then y is O_B
If x is M then y is O_M
If x is B then y is O_S.

The granules S, M, B, O_B, O_M, and O_S help us to grasp the input-output relationships. We interpret the models in Fig. 1 in the form of above rules with the granules. Thus the model in Fig. 1(a) with crisp and equidistantly allocated granules is interpretable.

Fig.1(b) and (c) show cases with Gaussian and triangular membership functions, respectively. Fig.1(d) is the case where triangular membership functions are allocated unevenly. Every model can be interpreted as the above three rules. But it becomes more and more difficult to make correspondence between the models and the three rules. The model in Fig.1(d) is the most difficult to interpret among the four models in Fig.1.

Assume that we have the following knowledge about the target system: "the non-linearity of the system becomes stronger with larger x". In this case, the interpretability of the models in Fig. 1 changes drastically from the above observation. We may think that the model in Fig. 1(d) is the most interpretable, because this model fits our knowledge most.

Interpretability of fuzzy models depends on prior knowledge. For quantitative analysis of interpretability, this paper limits the discussions in the following sections to the cases where no prior knowledge is available.

III. Fuzzy Model

A single-input single-output fuzzy model with simplified fuzzy inference[8] is used in this paper.

The output y of a fuzzy model with the input x is given by

$$y = \sum_{i=1}^{N_m} \mu_i(x) c_i \tag{1}$$

where $\mu_i(x)$ and c_i ($i = 1,\ldots, N_m$) are grade of membership in the antecedent part and singleton in the consequent part, respectively. N_m is the number of membership functions.

The following conditions are used to make the discussion about the conciseness simple:

(a) For all $x \in X$, membership functions $\mu_i(x)$ ($i = 1,\ldots, N_m$) satisfy

$$\sum_{i=1}^{N_m} \mu_i(x) = 1 \tag{2}$$

(b) Two membership functions overlap where $1 > \mu_i(x) > 0$ ($i = 1,\ldots, N_m$).

(c) Each membership function $\mu_i(x)$ is similar with respect to the center point $x = a$, in the sense that

$$\mu_{li}(x) = \mu_{ri}(1 - \frac{1-a}{a}x) \tag{3}$$

where $\mu_{li}(x)$ and $\mu_{ri}(x)$ are left/right hand side membership functions, respectively.

(d) All the fuzzy sets are convex.

IV. Conciseness of Fuzzy Model

The conciseness of fuzzy models is defined as the easiness for grasping the correspondence between the discrete fuzzy rules and the continuous values.

A. Definition of Conciseness

The conciseness of fuzzy models is defined in this paper as follows:

Definition 1(Conciseness of Fuzzy Model)
Fuzzy model A is more concise than fuzzy model B, if the membership functions in A are more uniformly distributed across the universe of discourse than the membership functions in B, and the shapes of membership functions in A are less fuzzy than in B.

B. De Luca and Termini's fuzzy entropy

De Luca and Termini [7] defined fuzzy entropy of fuzzy set A as

$$d(A) = \int_{x_1}^{x_2} \{-\mu_A(x)\ln\mu_A(x) - \mu_A(1-x)\ln\mu_A(1-x)\} \tag{4}$$

where $\mu_A(x)$ is the membership function of fuzzy set A. If $\mu_A(x) = 0.5$ for all x on the support of A, then the fuzzy entropy of fuzzy set A is the maximum.

This fuzzy entropy can distinguish the shapes of membership functions, i.e. triangular, Gaussian, etc., and coincides with a part of the definition of conciseness.

With the conditions (a) and (b) in Section III, De Luca and Termini's entropy is simplified. Assuming that two membership functions $\mu_A(x)$ and $\mu_B(x)$ are overlapping and for all $x \in [x_1, x_2]$, $\mu_A(x) + \mu_B(x) = 1$, then

$$d(A) = \int_{x_1}^{x_2} \{-\mu_A(x)\ln\mu_A(x)\}. \tag{5}$$

C. Measure for Deviation of Membership Function

De Luca and Termini's entropy cannot distinguish similar membership functions shown in Fig.2. Two membership functions A and C are similar in the sense that the membership function A coincides with C by extending the horizontal axis x of the left hand side membership function μ_{LA} and shrinking that of the right hand side membership function μ_{rA}.

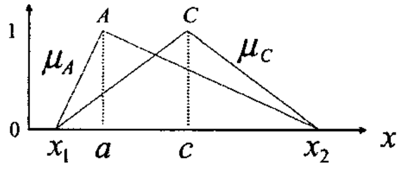


Fig. 2 Symmetrical/asymmetrical membership function

This paper defines a quantitative measure of deviation of a membership function from symmetry on the analogy of relative entropy. This is also a good candidate for the conciseness measure of fuzzy models. The membership function is assumed to satisfy the conditions (a) - (d) in Section III. This measure is defined by considering eq. (5).

Definition 2 (Measure for Deviation)
The measure for deviation of fuzzy set A from symmetry is given by

$$r(A)=\int_{x_1}^{x_2}\left(\mu_c \ln\frac{\mu_c}{\mu_A}\right)dx \qquad (6)$$

where x_1 and x_2 are the left and right points of the support of fuzzy set A, respectively; $\mu_A(x)$ is the membership function of fuzzy set A; $\mu_C(x)$ is the symmetrical membership function of fuzzy set C, which has the same support as that of fuzzy set A.

Fig. 2 illustrates an example of fuzzy set A and C.

D. New Measure

One way of combining the two measures is summation. By summing the fuzzy entropy $d(A)$ in eq.(5) and the measure for deviation of a membership function $r(A)$ in eq.(6), a new measure $dr(A)$ is obtained.

$$\begin{aligned} dr(A) &= d(A)+r(A) \\ &= -\int_{x_1}^{x_2}(\mu_C \ln \mu_A)dx. \end{aligned} \qquad (7)$$

E. Average Measure

Average measure dr_{avr} is introduced to evaluate the shapes and allocations of N_m fuzzy sets A_i (i = 1,..,, N_m) on the universe of discourse X on x-axis.

The average measure dr_{avr} is defined as

$$dr_{avr}=\frac{1}{N_m-2}\sum_{i=2}^{N_m-1} dr(A_i) \qquad (8)$$

where $dr(A)$ is the new measure in (7), which evaluates the shape and deviation of a membership function, N_m is the number of fuzzy sets A_i (i = 1,..., N_m) on the universe of discourse X on x-axis.

V. Numerical Results

This section describes results of a numerical experiment to demonstrate the feasibility of the average fuzzy entropy for evaluation of the conciseness of fuzzy models. For simplicity, the following single-input single-output function is used as a target function. Fig. 3 depicts this function.

$$f(x)=\begin{cases} 1-2x & (0\le x\le 1/2) \\ -4x^2+8x-3 & (1/2< x\le 1) \end{cases} \qquad (9)$$

Input-output pairs of data were generated using this function. The conditions (a) - (d) in Section III were imposed on the model. The accuracy of the obtained model was measured by mean squared error.

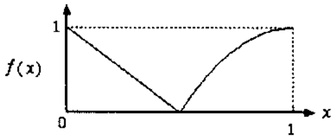


Fig. 3 Target function

To examine the relationships between the average measure dr_{avr} and the accuracy, 1000 fuzzy models were randomly generated and their values of average measure and accuracy were calculated. Fig. 4 shows the obtained results. Among them, the fuzzy models near the Pareto front are shown in Fig. 5.

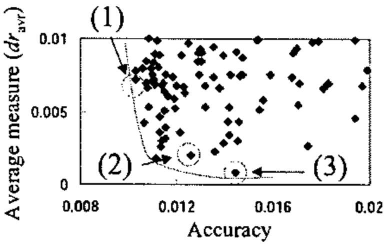


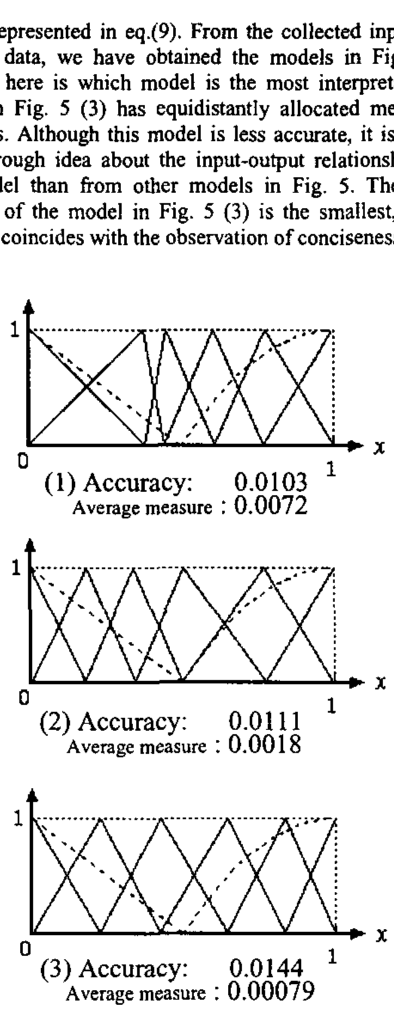
Fig. 4 Generated fuzzy models

In this paper, the shape of membership functions was fixed to triangular and the number of membership functions of a fuzzy model was set at 6. Each dot in Fig. 4 corresponds to a fuzzy model that has a unique allocation of membership functions. From Fig. 4, it is observed that the average measure and the accuracy are in conflict as indicated with the solid line.

Fig.5 (1)(2)(3) show the allocations of the membership functions of the fuzzy models (1)(2)(3), which were on the Pareto front in Fig.4 indicated by the circles. The less the average measure was, the more equidistant the allocation of membership functions was.

VI. Discussion

Assume that we have no prior knowledge about the target system represented in eq.(9). From the collected input-output pairs of data, we have obtained the models in Fig. 5. The question here is which model is the most interpretable. The model in Fig. 5 (3) has equidistantly allocated membership functions. Although this model is less accurate, it is easier to get the rough idea about the input-output relationships from this model than from other models in Fig. 5. The average measure of the model in Fig. 5 (3) is the smallest, and this measure coincides with the observation of conciseness.



(1) Accuracy: 0.0103
Average measure : 0.0072

(2) Accuracy: 0.0111
Average measure : 0.0018

(3) Accuracy: 0.0144
Average measure : 0.00079

Fig. 5 Allocation of membership functions of

obtained models on the Pareto front in Fig. 4

Next, assume that we have the knowledge about the target system *a priori*. This knowledge is expressed, for example, as "it is linear and decreasing on the left half of the universe of discourse, and is sharply rising up from the central point with increasing *x*." In this case, we may think the model in Fig. 5 (1) is the most interpretable among the models in this figure, because this model fits our prior knowledge most. On the other hand, the model in Fig. 5 (3) is the least interpretable now. This case shows that the interpretability of models depends on our knowledge.

VII. Conclusion

This paper discussed interpretability of fuzzy models by separating the cases with/without prior knowledge about target systems. Interpretability depends on the knowledge. For quantitative analysis of interpretability, this paper presented a new measure of conciseness of fuzzy models by focusing the cases where no prior knowledge was available. This paper defined the conciseness of fuzzy models, and quantified the conciseness by introducing De Luca and Termini's fuzzy entropy, and also defined a new measure of the deviation of a membership function from symmetry. By combining the new measure and De Luca and Termini's measure the new measure for the conciseness was derived. Based on the new measure, the average measure was defined to evaluate the shapes and the deviation of allocation of membership functions of fuzzy models. Experimental results showed that the new measure coincide with the observation of conciseness, and it was in conflict with the accuracy of fuzzy models. The results also showed that the least concise model was the most interpretable in the case where we had no prior knowledge. The interpretability was shown to be changed with *a priori* knowledge about the target system.

References

[1] S. Matsushita and T. Furuhashi, et al., "Selection of Input Variables Using Genetic Algorithm for Hierarchical Fuzzy Modeling," Proc. of 1996 First Asia-Pacific Conference on Simulated Evolution and Learning, pp.106-113, 1996.

[2] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," 2nd International Symposium on Information Theory, pp.267-281, 1973.

[3] H. Nomura, S. Araki, I. Hayashi and N. Wakami, "A Learning Method of Fuzzy Reasoning by Delta Rule," Proc. of Intelligent System Symposium, pp.25-30, 1992.

[4] M. Setnes and R. Babuska and H.B.Verbruggen, "Rule-Based Modeling: Precision and Transparency," IEEE Trans. Syst., Man, Cybern., pt.C, Vol. 28, No. 1, pp.165-169, Feb. 1998.

[5] J. Valente de Oliveira, "Semantic Constraints for Membership Function Optimization," IEEE Trans. Syst., Man, Cybern., pt.A, Vol. 23, No. 1, pp.128-138, Jan., 1999.

[6] M. Setnes and H. Roubos, "GA-Fuzzy Modeling and Classification: Complexity and Performance," IEEE Trans. Fuzzy Syst., Vol. 8, No. 5, pp.509-522, Oct. 2000.

[7] A. De Luca and S. Termini, "A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory," Information and Control, Vol. 20, pp.301-312, 1972.

[8] M. Mizumoto, "Fuzzy Control Under Various Approximate Reasoning Methods," Proc. of Second IFSA Congress, pp.143-146, 1987.