Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Fuzzy Sets and Systems 142 (2004) 243 – 265
FUZZY sets and systems
www.elsevier.com/locate/fss

Rule-base self-generation and simplification for data-driven fuzzy models

Min-You Chen*, D.A. Linkens

Department of Automatic Control and Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Received 14 October 2002; received in revised form 31 March 2003; accepted 4 April 2003

Abstract

Data-driven fuzzy modeling has been used in a wide variety of applications. However, in fuzzy rule-based models acquired from numerical data, redundancy often exists in the form of redundant rules or similar fuzzy sets. This results in unnecessary structural complexity and decreases the interpretability of the system. In this paper, a rule-base self-extraction and simplification method is proposed to establish interpretable fuzzy models from numerical data. A fuzzy clustering technique associated with the proposed fuzzy partition validity index is used to extract the initial fuzzy rule-base and find out the optimal number of fuzzy rules. To reduce the complexity of fuzzy models while keeping good model accuracy, some approximate similarity measures are presented and a parameter fine-tuning mechanism is introduced to improve the accuracy of the simplified model. Using the proposed similarity measures, the redundant fuzzy rules are removed and similar fuzzy sets are merged to create a common fuzzy set in the rule base. The simplified rule base is computationally efficient and linguistically interpretable. The approach has been successfully applied to fuzzy models of non-linear function approximation, dynamical system modeling and mechanical property prediction for hot-rolled steels.

Keywords: Fuzzy modeling; Fuzzy clustering; Rule-base simplification; Similarity analysis

1. Introduction

Data-driven fuzzy modeling is finding a growing number of significant applications in a wide variety of fields ranging from pattern recognition, data mining, classification, prediction, non-linear system approximation, and process control [4,6,8–10,13,17,18,20,21]. Primary advantages of fuzzy modeling include the facility for the explicit knowledge representation in the form of if-then rules, the mechanism of human-like reasoning in linguistic terms, and the ability to approximate complicated

* Corresponding author. Tel.: +44-1142225611; fax: +44-1142731729.
E-mail addresses: minyou.chen@shef.ac.uk (M.-Y. Chen), d.linkens@shef.ac.uk (D.A. Linkens).

0165-0114/$ - see front matter
doi:10.1016/S0165-0114(03)00160-X

non-linear functions with simpler models. However, the rule base automatically generated from data may not be interpretable because redundancy in the form of similar fuzzy sets usually exists in data-driven fuzzy models. This results in unnecessary complexity and poor transparency of the rule-based model. To improve the interpretability of fuzzy models, several methods have been proposed. Some of them focused on the tradeoff between numerical accuracy and linguistic interpretability [11,15,19]. In these methods, a formulation of some constraints is imposed in the optimization of the membership functions to guarantee semantic integrity. The fuzzy rules are constructed within a framework of linguistic integrity to guarantee its interpretability while maintaining good fitness of input/output data. The main drawback of this kind of methods is the exponential growth of the rule base as the number of inputs increases [19]. The others emphasize on the trade-off between model accuracy and simplicity [3,7,14,24]. These methods tried to simplify the acquired data-driven fuzzy models to improve the interpretability while maintaining model accuracy. In practical applications, the trade-off between model accuracy and simplicity is a fundamental of fuzzy modelling, which is also the focus of this paper. In recent years, several approaches to fuzzy model generation and simplification have been proposed. Chao and Chen [3] proposed a fuzzy rule-base simplification method based on similarity analysis. Some fuzzy similarity measures based on triangular membership function were proposed to eliminate redundant fuzzy rules and combine similar input linguistic terms. Setnes et al. [14] presented a rule-base simplification approach using set-theoretic similarity measures to reduce the number of fuzzy sets in the models. Yen and Wang [24] introduced several orthogonal transformation-based methods to select a set of important fuzzy rules from a given rule-base. In these methods, a pre-determined number of fuzzy rules is required to build initial fuzzy model. More recently, Jin [7] proposed a fuzzy modeling approach for high-dimensional systems. A distance-based similarity measure was adopted to check the similarity of fuzzy sets to remove the redundancy in the rule-base, and a regularized learning was introduced to improve the interpretability of fuzzy models. However, the rule generation approach based on output extreme proposed in [7] could generate a relatively large rule-base and may not result in an optimal model structure. In this paper, a more general and effective fuzzy rule-base extraction and simplification method is presented, which includes fuzzy clustering with partition validation, gradient-descent based parameter optimization and approximate similarity analysis based rule-base simplification. The number of fuzzy rules is determined automatically by the fuzzy clustering procedure associated with the proposed partition validity index. By using the fuzzy similarity measure, we derive simple approximate equations for calculating the degree of similarity of two fuzzy sets, both with symmetric membership functions. The approximate similarity measures for eliminating redundant fuzzy sets are presented and applied to combine similar linguistic terms into a single linguistic term to reduce the complexity of the fuzzy models. Thus we attempt to produce a simple and interpretable fuzzy inference system with satisfactory accuracy, which is more practical and useful in industrial applications.

2. Generating fuzzy models from numerical data

Fuzzy modeling can be interpreted as a qualitative modeling scheme which describes system behavior using fuzzy quantities [17], i.e. fuzzy sets or fuzzy numbers. Generally, the data-driven fuzzy modeling problem can be formulated as follows:

Given the n input/output patterns $P(\boldsymbol{x}, y)$ and the specified model error $\varepsilon > 0$, obtain the minimal number p of fuzzy rules and optimal parameters, including membership function parameters θ in the antecedent part and linear weights $\boldsymbol{w}$ in the consequent part of the rules, for the fuzzy model $F(\theta, \boldsymbol{w}, p)$ such that the error function $E = \|y - F\|$ satisfies the inequality $E(\theta, \boldsymbol{w}) < \varepsilon$.

Based on a collection of s-dimensional data points $\{\boldsymbol{P}_1, \boldsymbol{P}_2, \ldots, \boldsymbol{P}_n\}$, a multi-input and single-output (MISO) fuzzy model is represented as a collection of fuzzy rules in the following form:

$$R_i: \quad \text{IF } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \quad \ldots \quad \text{and } x_s \text{ is } A_{is} \text{ THEN } y_i = z_i(\boldsymbol{x}),$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_s) \in U_1 \times U_2 \times \cdots \times U_s$ are linguistic variables, A_{ij} are fuzzy sets of the universes of discourse $U_j \in R$ ($j = 1, 2, \ldots, s$), R_i represents the ith rule, $i = 1, 2, \ldots, p$, and $y_i \in V$ is the output of the ith rule. Typically, $z_i(\boldsymbol{x})$ takes the following forms: singleton, i.e. $z_i = b_i$, which can be represented as Mamdani model, or linear function, i.e. $z_i(\boldsymbol{x}) = b_{i0} + \sum_{j=1}^{s} b_{ij} x_j$, which is Takagi–Sugeno (TS) model. In this paper we are concerned with the identification of TS models, since a Mamdani model can be considered as a zero order TS model under certain conditions.

Fuzzy logic systems with center of average defuzzification, product-inference-rule and singleton fuzzification are of the following form:

$$y = \sum_{i=1}^{p} z_i \left[\prod_{j=1}^{s} u_{ij}(x_j) \right] \Bigg/ \sum_{i=1}^{p} \prod_{j=1}^{s} u_{ij}(x_j) , \tag{1}$$

where $u_{ij}(x_j)$ denotes the membership function of x_j belonging to the ith rule. Very commonly, a radial basis function, especially the Gaussian function is chosen as the membership function, i.e.

$$u_{ij}(x_j) = \exp\left(-\frac{(x_j - a_{ij})^2}{\sigma_{ij}^2} \right), \tag{2}$$

where a_{ij} and σ_{ij} are center and width of the jth membership function in the ith rule.

Thus, Eq. (1) can be rewritten as

$$y = \sum_{i=1}^{p} z_i q_i(\boldsymbol{x}), \tag{3}$$

where $q_i(\boldsymbol{x}) = m_i(\boldsymbol{x}) / \sum_{i=1}^{p} m_i(\boldsymbol{x})$, and $m_i(\boldsymbol{x}) = \prod_{j=1}^{s} u_{ij}(x_j)$ represents the matching degree of the current input $\boldsymbol{x}$ to the ith fuzzy rule.

Our aim is to develop an automatic rule generating mechanism, without any assumption about the structure of the data, which is capable of (1) generating a rule base automatically from numeric data, (2) finding the optimal number of the rules, and (3) simplifying the obtained fuzzy model to enhance the model interpretability while maintaining satisfactory accuracy. It is noted that in the case of rule extraction from data, an effective data partition in input–output space can lead to reducing the number of rules and thus improving the computational efficiency and interpretability of the fuzzy models. The above objectives can be achieved by incorporating fuzzy c-means (FCM) clustering associated with a new partition validity index and the approximate similarity analysis.

2.1. Rule-base self-extraction

It is generally acknowledged that classes or clusters of the data which have similar geometrical location should be formed. Fuzzy clustering is a well recognized paradigm to generate the initial fuzzy model. Numerous clustering algorithms have been developed. The most widely used algorithm is the FCM due to its efficacy and simplicity. However, the number c of clusters must be pre-determined. The FCM algorithm partitions a collection of n data points ($X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\}$) into c fuzzy clusters such that the following objective function is minimized:

$$J_m = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m(x) \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2, \quad 1 < m < \infty, \tag{4}$$

where m is a fuzzy exponent, $\boldsymbol{v}_i$ is the prototype of the ith cluster generated by fuzzy clustering, $u_{ik} \in [0,1]$ is the membership degree of the kth data belonging to the ith cluster represented by $\boldsymbol{v}_i$, $u_{ik} \in U$, U is a $c \times n$ fuzzy partition matrix which satisfies the constraints:

$$0 < \sum_{k=1}^{n} u_{ik} < n \quad \text{for } i = 1, 2, \dots, c; \quad \text{and} \quad \sum_{i=1}^{c} u_{ik} = 1 \quad \text{for } k = 1, 2, \dots, n.$$

Cluster validity is the problem of finding the best value for c subject to minimization of J_m. Since J_m monotonically decreases with c, an effective criterion for evaluating the partition quality is required. Many cluster-validity criteria have been proposed to measure the effectiveness of the clustering. The first fuzzy cluster-validity criteria associated with FCM introduced by Bezdek are the *partition coefficient* (*PC*) and the *partition entropy* (*PE*) [1,2]. Their main advantage is their simplicity but the main disadvantage is their monotonic tendency with c [12]. Fukayama and Sugno [5] and Xie and Beni [23] introduced new fuzzy validity criteria for evaluating fuzzy c-partitions, which are commonly used as fuzzy cluster validity measures. They combine, with a unique function, the properties of the fuzzy membership degrees and the structure of the data. These criteria provide useful tools for cluster validation, each of which has developed its own set of partially successful validation schemes although they lose their ability to validate partitions from FCM for large m [12]. A good validity index must take into account both compactness and separation of clusters in its partitioning. In this paper, a simple and effective fuzzy partition measure is proposed as a cluster validity criterion associated with the FCM algorithm, which is defined as follows:

$$V_\mathrm{p}(U, c) = \frac{1}{n} \sum_{k=1}^{n} \max_i (u_{ik}) - \frac{1}{K} \sum_{i=1}^{c-1} \sum_{j=i+1}^{c} \left[\frac{1}{n} \sum_{k=1}^{n} \min(u_{ik}, u_{jk}) \right], \quad \text{where } K = \sum_{i=1}^{c-1} i. \tag{5}$$

It can be seen that the cluster validity measure V_p is composed of two terms. The first one reflects the compactness within a cluster. The closer the kth pattern $\boldsymbol{x}_k$ is to a fuzzy cluster centre, the closer the maximum membership degrees $\max_i(u_{ik})$ is to the value 1. Hence, the fuzzy set $\max_i(u_{ik})$ is considered as a good indicator of the clustering quality for each pattern $\boldsymbol{x}_k$. This quality indicates how closely the objects are assigned to the fuzzy cluster centres. Thus, a large value of the first item indicates that the data patterns are well classified. On the other hand, the second term indicates the separation between clusters. Here, the intersection of two fuzzy sets is used to evaluate a fuzzy separation between clusters $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$. In fact, if $\boldsymbol{x}_k$ is close to the fuzzy cluster centre of $\boldsymbol{v}_i$,

$\min(u_{ik}, u_{jk})$ comes close to 0, and consequently the fuzzy sets U_i and U_j are clearly separated. On the other hand, if $\min(u_{ik}, u_{jk})$ is close to $1/c$, $\boldsymbol{x}_k$ belongs to all clusters with equal membership degree and the fuzziest partition is obtained. The new validity V_p criterion combines information on fuzzy compactness and separation. It tends to indicate a good cohesion within clusters and a small overlap between pairs of clusters. Thus, the number of clusters c corresponding to the maximum value of V_p indicates the optimal number of clusters. In contrast to the indexes $PC \in [1/c, 1]$ and $PE \in [0, \log_\mathrm{a} c]$, both lower and upper bound of $V_\mathrm{p} \in [0, 1]$ are independent of c, which overcomes the main shortcoming of PC and PE. According to experimental results, the proposed validity index V_p works very well in the range of $m \in [1.5, 5]$, which is very usual in practice ($m = 2$ is so far the most common choice). It is noted that the proposed index V_p is computationally simple and could be used as an alternative in existing fuzzy partition validity criteria. A comparative study between different validity indices is given in Section 4.

The FCM algorithm attempts to classify the given set of data vectors into a certain number of clusters by searching for local minima of J_m. The procedure of the fuzzy clustering algorithm associated with the validity measure (5) is carried out in the product space of input–output variables through an iterative optimization of J_m according to the following steps:

Step 1: Choose the maximum cluster number $c_{\max}$ (heuristically, $c_{\max} \leqslant \sqrt{n}$), iteration limit T, weighting exponent m, and termination criterion $\varepsilon > 0$.

Step 2: With $c = 2, 3, \ldots, c_{\max}$; initialize the position of cluster centres: $\boldsymbol{V}_0 = (\boldsymbol{v}_{10}, \boldsymbol{v}_{20}, \ldots, \boldsymbol{v}_{c0})$;

Step 3: With the iteration number $t = 1, 2, \ldots, \mathrm{T}$;

$$\text{calculate } u_{ik,t} = 1 \Bigg/ \sum_{j=1}^{c} (d_{ik}/d_{jk})^{2/(m-1)}, \tag{6}$$

where $d_{ik} = \|\boldsymbol{x}_k - \boldsymbol{v}_i\|$, $i = 1, 2, \ldots, c$; $k = 1, 2, \ldots, n$;

$$\text{calculate } \boldsymbol{v}_{i,t} = \sum_{k=1}^{n} (u_{ik,t})^m x_k \Bigg/ \sum_{k=1}^{n} (u_{ik,t})^m . \tag{7}$$

If $\|\boldsymbol{V}_t - \boldsymbol{V}_{t-1}\| < \varepsilon$, go to next step, otherwise repeat step 3.

Step 4: Calculate $V_\mathrm{p}(\mathrm{U}, c)$ by (5); if $c < c_{\max}$, repeat from Step 2. Otherwise, stop the program and set the optimal cluster number $c = c_\mathrm{m}$, where c_m meets the following condition:

$$V_\mathrm{p}(U, c_\mathrm{m}) = \max\{V_\mathrm{p}(U, c)\}, \quad c = 2, 3, \ldots, c_{\max}.$$

After cluster validation, both the number of rules and the prototypes of the clusters $\boldsymbol{v}_i = (v_{i1}, v_{i2}, \ldots, v_{is}, v_{i,s+1})$, are obtained, where $i = 1, 2, \ldots, c$. Let $\boldsymbol{a}_i = (a_{i1}, a_{i2}, \ldots, a_{is}) = (v_{i1}, v_{i2}, \ldots, v_{is})$, $z_i = v_{i,s+1}$, then the vector $\boldsymbol{a}_i$ denotes the prototype of the ith fuzzy partition in the input space, and it can also be viewed as the center values of Gaussian membership functions in the antecedent of the ith rule, while z_i is the prototype of the ith fuzzy partition in the output space, and denotes the fuzzy output value in the consequent part of the ith rule.

Therefore, the rule-base which is composed of c fuzzy rules can be represented as

$$R_i: \quad \text{IF } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \ldots \text{ and } x_s \text{ is } A_{is} \text{ THEN } y \text{ is } z_i,$$

where R_i denotes the ith rule, A_{ij} is the fuzzy set defined by the Gaussian membership function; and $z_i = b_i$ or $z_i = \sum_{j=0}^{s} b_{ij}x_j$, is the ith rule output with respect to a Mamdani model or a TS model.

2.2. Parameter estimation

When an initial fuzzy model is constructed in the process of rule-base generation, a parameter learning procedure is successively applied to obtain a more precise fuzzy model in the process of parameter identification. There are several methods for training the fuzzy model, that is, to learn the optimal membership function parameters a_{ij}, σ_{ij} and linear weights b_{ij}. Here, we adopt the gradient-descent-based approach to optimize the parameters a_{ij}, σ_{ij} and b_{ij} in combination within the performance index of *mean square error* (MSE). Using gradient-descent algorithms, the parameter learning algorithms can be derived as

$$\Delta b_{ij} = \eta(y_{dk} - y_k)x_j q_i(\boldsymbol{x}), \tag{8}$$

$$\Delta a_{ij} = \eta(y_{dk} - y_k)\frac{(x_{jk} - a_{ij})}{\sigma_{ij}^2}(z_i - y_k)q_i(\boldsymbol{x}), \tag{9}$$

$$\Delta \sigma_{ij} = -\eta(y_{dk} - y_k)\frac{(x_{jk} - a_{ij})^2}{\sigma_{ij}^3}(z_i - y_k)q_i(\boldsymbol{x}), \tag{10}$$

where η is the learning rate, y_{dk}, and y_k are desired output and model output, respectively.

3. Approximate similarity measures

After parameter learning, the optimal fuzzy rule-base is not yet finally constructed. The obtained fuzzy model may exhibit redundancy in terms of highly overlapping membership functions. To acquire an efficient and transparent fuzzy model, elimination of redundancy and making the fuzzy model as simple as possible is necessary. Some similarity measures have been introduced to simplify fuzzy models [3,7,9,14]. Generally, there are two types of similarity analysis methods: fuzzy set-theory based similarity measure and geometric graph based similarity measure, for computing the similar degree of two fuzzy sets. The former calculates the similarity based on the size (or cardinality) of the fuzzy sets, which is usually not easy and straightforward for non-linear membership functions. Recently, a simple similarity measure based on triangular membership functions has been proposed and used to calculate the similarity degree of two fuzzy sets with Gaussian membership functions [3,9]. This section presents approximate fuzzy similarity measures which can be used to any symmetric fuzzy sets.

Assume A and B are two fuzzy sets, the similarity of fuzzy sets A and B being defined as

$$S_{AB} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}, \tag{11}$$

where $\cap$ and $\cup$ denote intersection and union of A and B, respectively. $|\cdot|$ denotes the size of a fuzzy set. Clearly, computation of two fuzzy sets requires calculating the size of intersection of the two fuzzy sets. For Gaussian and bell-shaped functions, it is complex to compute the size of the

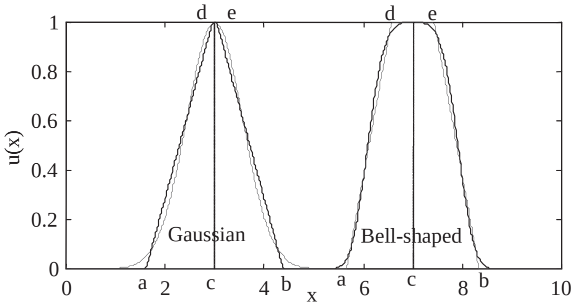

Fig. 1. Membership function approximation.

intersection because of the non-linear shape of the functions. To make the computation of (11) feasible, Chao et al. [3] and Lin and Lee [9] proposed approximate similarity analysis approaches, which use triangular functions as the tent function to calculate the similarity of two Gaussian membership functions. However, it is difficult to use a triangular function to approximate Generalized Gaussian or bell-shaped functions, which are commonly used in fuzzy systems. It is found that a trapezoidal function can approximate a radial basis function (including Gaussian and bell-shaped functions) very well, as shown in Fig. 1, where a, b and d, e denote the boundary points of a trapezoid, and c denotes the central point. Also, a triangular MF can be viewed as the special case of a trapezoidal MF when its top width $w_t = 0$, i.e. $d = e = c$. Hence, we can use a trapezoidal function as the tent function to calculate similarity of two symmetric membership functions, including triangular, Gaussian, Generalized Gaussian, and bell-shaped functions. The problem is how to generate a trapezoidal function that can subsequently approximate any radial basis function well.

3.1. Generating trapezoidal membership functions from radial basis functions

To determine a trapezoidal function, $T(x) = \max\{\min\{(x-a)/(d-a), 1, (b-x)/(b-e)\}, 0\}$, which can approximate a generalized Gaussian function $G(x) = \exp\{-[|x-c|/\sigma]^m\}$, or more generally, fit any symmetric membership functions, we introduce the α-cut of a fuzzy set defined as follows:

The α-cut of a fuzzy set A, denoted as A_α, is the crisp set comprised of all the elements x of a universe of discourse X for which the membership function of A is greater than or equal to α, i.e.

$$A_\alpha = \{x \in X \mid \mu(x) \geqslant \alpha\}, \tag{12}$$

where α is a parameter in the range $0 < \alpha \leqslant 1$; the vertical bar "|" is shorthand for "such that". So, the α-cut (or α-level) set of a fuzzy set A is a closed interval of R. Assume a normal fuzzy set A is represented by a generalized Gaussian membership function $G(x)$, which can be approximated by a trapezoidal membership function $T(x)$, as shown in Fig. 2. To identify the parameters a, b, d, e in $T(x)$, on the basis of $G(x)$, we introduce two special α-cut sets of A, the bottom α-cut A_{α_0} and the top α-cut A_{α_1}, where $\alpha_0 = 0.05$, $\alpha_1 = 0.95$, as shown in Fig. 2. Thus, the parameters a, b, d, e can be decided on the basis of $G(x)$ via the α-cuts cut A_{α_0} and A_{α_1}, which are represented as

$$A_{\alpha_0} = [a, b], \quad A_{\alpha_1} = [d, e].$$

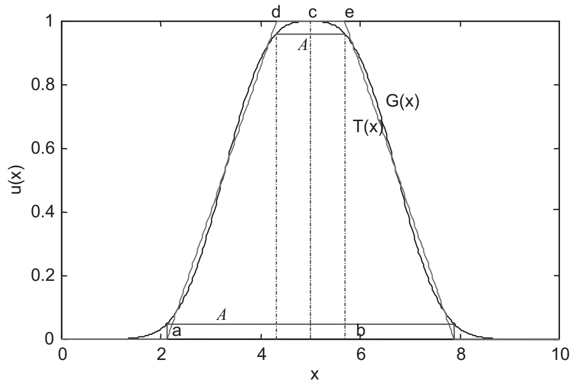

Fig. 2. α-cut of a fuzzy set.

Once the boundary points a, b, d, and e are obtained, the top width w_t and bottom width w_b can be derived as: $w_t = (e - d)/2$, $w_b = (b - a)/2$. The centre of the trapezoidal function is given by

$$c = (d + e)/2 \quad \text{or} \quad c = (a + b)/2.$$

3.2. Approximate similarity measures for fuzzy sets

Based on trapezoidal membership functions, the similarity measure of two fuzzy sets can be considered for four different cases. The fuzzy sets are denoted by A_1 and A_2, with the corresponding centres c_1 and c_2, and boundary points a_i, b_i, d_i, e_i. $(i = 1, 2)$, where $c_i = (e_i + d_i)/2$.

Assume $c_2 > c_1$ in cases (i)–(iv).

Case (i) ($a_1 < a_2$, $b_1 > b_2$, & $d_1 \leqslant d_2$, $e_2 \leqslant e_1$). In this case, the fuzzy set A_2 is included in A_1, i.e. $A_2 \subset A_1$, as shown in Fig. 3. The similarity of fuzzy sets A_1 and A_2 can be calculated simply by

$$S = \frac{|A_2|}{|A_1|} = \frac{w_2}{w_1}, \quad \text{i.e. } S = \frac{b_2 - a_2 + e_2 - d_2}{b_1 - a_1 + e_1 - d_1}, \tag{13}$$

where $w_1 = w_{b1} + w_{t1}$, $w_2 = w_{b2} + w_{t2}$.

From (13) we can see that the degree of similarity of A_1 and A_2 is just the ratio of w_2 to w_1.

Case (ii) ($|w_{b1} - w_{b2}| \leqslant c_1 - c_2 \leqslant w_{b1} + w_{b2}$). In this case, there are two different overlapping situations, as shown in Figs. 4(a) and (b). In situation (a), there is no top width overlap between A_1 and A_2. The similarity of A_1 and A_2 can be derived as

$$S = h \Bigg/ \left(\frac{w_1 + w_2}{b_1 - a_2} - h \right), \tag{14}$$

where

$$h = \frac{b_1 - a_2}{b_1 - e_1 + b_2 - e_2}.$$

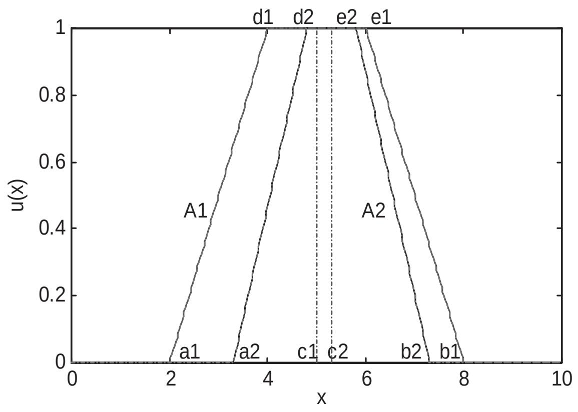

Fig. 3. Similarity of two fuzzy sets for case 1.

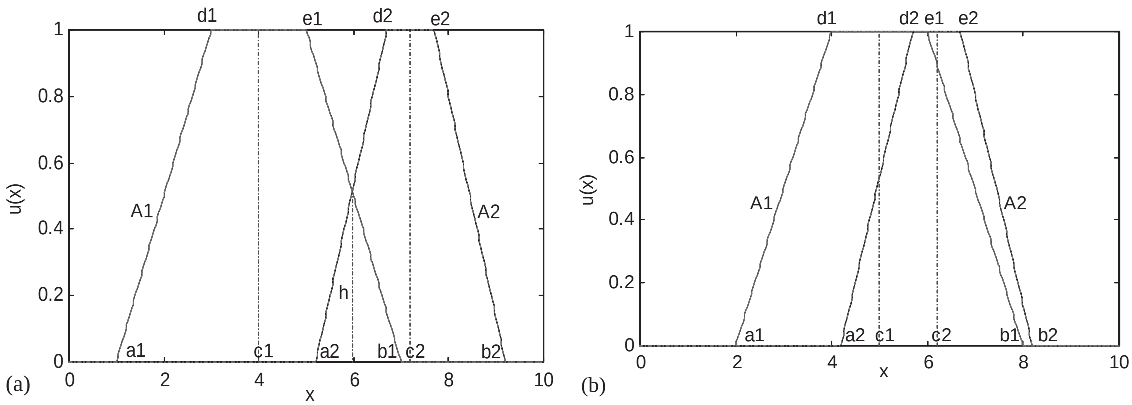

Fig. 4. Similarity of two fuzzy sets for case 2.

In case (b), there is top width overlap between A_1 and A_2. The similarity of A_1 and A_2 can be easily obtained as follows:

$$S = \frac{e_1 - d_2 + b_1 - a_2}{b_2 - a_1 + e_2 - d_1}. \tag{15}$$

Case (iii) ($c_1 - c_2 \leqslant |w_{b1} - w_{b2}|$). Again, this in case, we should consider two different overlap cases as shown in Fig. 5(a) and (b). We can easily derive

$$h_1 = \frac{b_1 - a_2}{(b_1 - e_1) + (b_2 - e_2)}.$$

For $w_{b1} > w_{b2}$,

$$h_2 = \frac{b_1 - b_2}{(b_1 - e_1) - (b_2 - e_2)},$$

where $l_1 = h_1(b_2 - e_2)$; $l_2 = h_2(b_2 - e_2)$; $l_3 = (b_2 - a_2) - (l_1 + l_2)$.

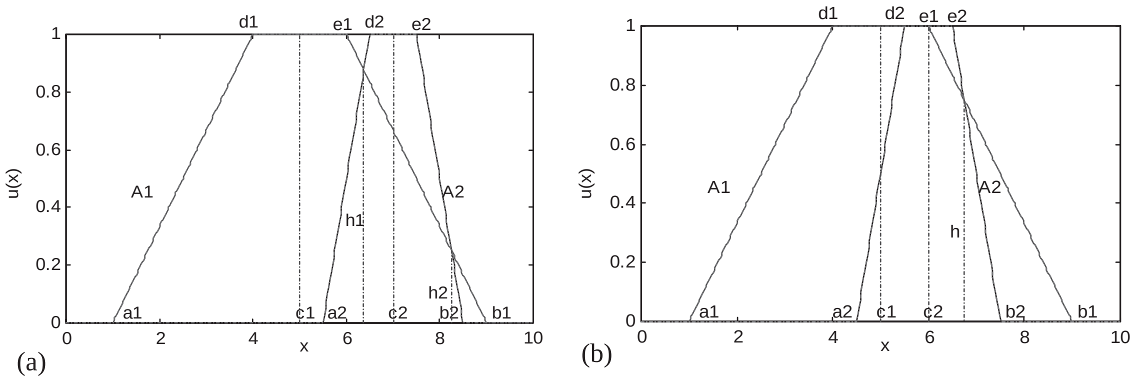

Fig. 5. Similarity of two fuzzy sets for case 3.

For $w_{b1} \leqslant w_{b2}$,

$$h_2 = \frac{a_2 - a_1}{(b_1 - e_1) - (b_2 - e_2)},$$

where, $l_1 = h_1(b_1 - e_1)$; $l_2 = h_2(b_1 - e_1)$; $l_3 = (b_1 - a_1) - (l_1 + l_2)$; $h_3 = h_1 + h_2$.

The similarity of A_1 and A_2 is given as follows:

$$S = \frac{l_1 h_1 + l_2 h_2 + l_3 h_3}{2(w_1 + w_2) - (l_1 h_1 + l_2 h_2 + l_3 h_3)}. \tag{16}$$

Overlap: $c_1 - c_2 <= w_{t1} + w_{t2}$

For $w_{b1} > w_{b2}$:

$$h = \frac{b_1 - b_2}{(b_1 - e_1) - (b_2 - e_2)}, \quad l_1 = h(b_2 - e_2), \quad l_2 = b_2 - e_1 - l_1,$$

$$l_3 = e_1 - d_2, \quad l_4 = b_2 - e_2.$$

For $w_{b1} \leqslant w_{b2}$:

$$h = \frac{a_2 - a_1}{(b_1 - e_1) - (b_2 - e_2)}, \quad l_1 = h(b_1 - e_1), \quad l_2 = d_2 - a_1 - l_1,$$

$$l_3 = e_2 - d_1, \quad l_4 = b_1 - e_1.$$

Thus we obtain $H = l_1 h + l_2(h + 1) + 2l_3 + l_4$, and

$$S = \frac{H}{2(w_1 + w_2) - H}. \tag{17}$$

Case (iv) ($b_1 \leqslant a_2$). In this case, there is no intersection between Membership functions of A_1 and A_2, as shown in Fig. 6, thus $|A_1 \cap A_2| = 0$, and $S(A_1, A_2) = 0$.

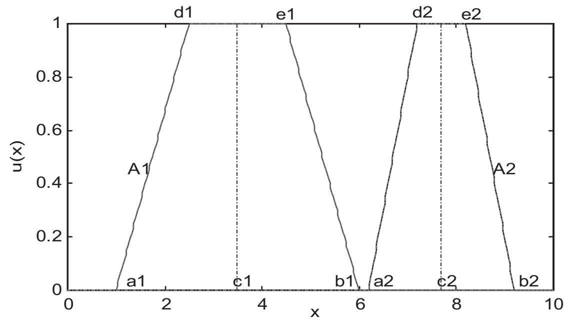

Fig. 6. Similarity of two fuzzy sets for case 4.

4. Simplification of fuzzy models

Based on the obtained fuzzy rule-base and the approximate similarity measures, a fuzzy model simplification approach for minimizing the number of fuzzy sets on the universe of discourses of each input variable and fuzzy rules is presented. The procedure of model simplification is shown in Fig. 7. It can be seen that the model simplification procedure consists of similarity-analysis-based fuzzy sets pruning and rule reduction. The implementation of the model simplification procedure is presented in the following subsections.

4.1. Reducing the number of fuzzy sets for each input variable

(1) *Redundant fuzzy sets removing*. For each fuzzy set A_{ij} calculate the similarity $S(A_{ij}, U_j)$, where $\mu_{U_j}(x_j) = 1$, $\forall x_j \in U_j$.

If $S(A_{ij}, U_j) > \lambda_r$, then remove A_{ij} from the antecedent of rule R_i; where $\lambda_r \in (0, 1)$ is the threshold for removing fuzzy sets similar to the universal set. Usually, we set $\lambda_r = 0.9$.

(2) *Similar fuzzy sets combination*. We can use different fuzzy similarity measures to check the degree of similarity of two fuzzy sets A_{ij} and A_{kj} in different cases. For symmetric continuous membership functions, the similarity between the fuzzy sets can be calculated by Eqs. (13)–(17). Otherwise, calculate the similarity by

$$S(A_{ij}, A_{kj}) = \frac{\sum_{l=1}^{L} \min\{u_{ij}(x_{jl}), u_{kj}(x_{jl})\}}{\sum_{l=1}^{L} \max\{u_{ij}(x_{jl}), u_{kj}(x_{jl})\}}, \quad j = 1, \ldots, s; \;\; i, k = 1, \ldots, c, \;\; i \neq k, \tag{18}$$

where L is the number of sampling data. If $S(A_{ij}, A_{kj}) > \lambda_{\mathrm{m}}$, then merge the two fuzzy sets A_{ij}, A_{kj} into one new fuzzy set A_{pj}, where $\lambda_{\mathrm{m}} \in (0, 1)$ is the threshold for merging fuzzy sets that are similar to one another. The center and the width of the new fuzzy set A_{pj} can be obtained simply by average values of the fuzzy sets A_{ij}, A_{kj}, i.e. $c_{pj} = (c_{ij} + c_{kj})/2$, $w_{pj} = (w_{ij} + w_{kj})/2$. It should be pointed out that the threshold λ_{m} influence the model performance significantly. Smaller λ_{m} results in more sets merge thus generates simpler fuzzy model, but usually lower model accuracy. We suggest $\lambda_{\mathrm{m}} = 0.6$ –0.85. It is important to keep balance between model simplicity and accuracy.

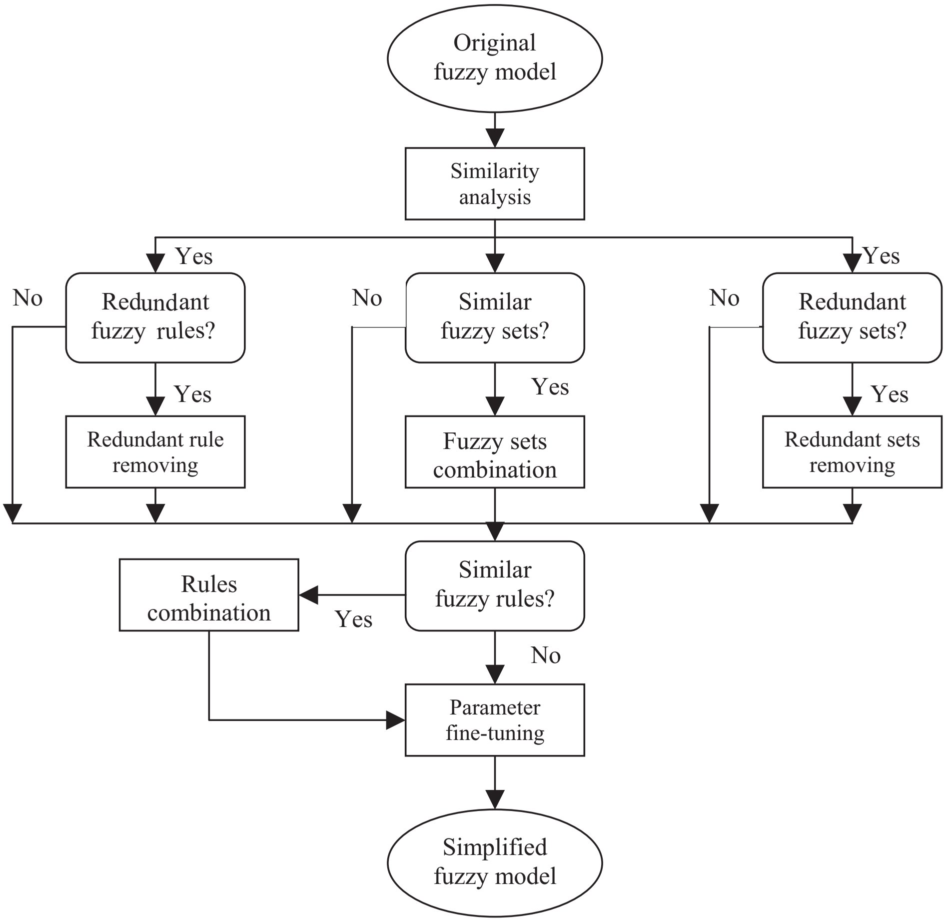

Fig. 7. Schematic diagram of fuzzy model simplification.

4.2. Reducing the number of fuzzy rules

If the redundancy in the model is high, removing the redundant rules with little influence and combining similar fuzzy rules into an equivalent fuzzy rule could result in further simplification of the fuzzy models. The former is referred to as rule elimination; the latter as rule combination.

(1) *Rule elimination*. If a fuzzy membership function is always near zero over its own universe of discourse, i.e. $\mu_{ji}(x_i) \to 0$ for $x_i \in U_i$, then remove the rule with this membership function because the output of this rule is always near zero.

(2) *Rule combination*. To decide whether two fuzzy rules are similar enough for combination, we only need to evaluate the similarity of antecedent parts of the rules. Two fuzzy rules with very similar antecedents but different consequents usually indicate that the two rules conflict each other. Therefore, we should either combine those rules into one new rule or delete one of them.

To calculate the degree of similarity of the antecedent of two fuzzy rules, we should check the similarity degree of every fuzzy set pair. With the jth fuzzy rule, the corresponding preconditions are $A_{j1}, A_{j2}, \ldots, A_{js}$. Similarly, the corresponding antecedent of the kth rule are $A_{k1}, A_{k2}, \ldots, A_{ks}$. Thus, the similarity measure of the antecedent can be characterized as follows:

$$S_{\rm a}(A_j, A_k) = \min_i \{S(A_{ji}, A_{ki})\}; \quad i = 1, 2, \ldots, s.$$

Once $S_{\rm a}(A_j, A_k)$ reaches a reference value $\gamma_{\rm a}$, then all of these fuzzy set pairs are considered to be very similar. Thus, the two rules can be combined into one new fuzzy rule $R_{\rm new}$. The antecedent of $R_{\rm new}$ can be directly obtained by using the combination methods of fuzzy sets presented in the previous subsection. On the other hand, the consequent of $R_{\rm new}$ can be simply chosen as $z_{\rm new} = (z_j, +z_k)/2$.

4.3. Parameter fine-tuning

After similarity analysis and rule-base pruning, the obtained model is structurally simpler and interpretably more tractable. However, this model is less accurate than the originally generated model. To improve the accuracy of the simplified model, a parameter fine-tuning procedure is necessary. In this work, we have introduced a parameter fine-tuning mechanism for the simplified fuzzy model using the same gradient-descent algorithms presented in Eqs. (8)–(10). Through fine-tuning, the loss of accuracy of the simplified model can be reduced to a minimum.

5. Illustrative examples

In order to demonstrate the validity of our method, different kinds of non-linear system modeling examples are presented in this section. These examples cover the range of fuzzy clustering, non-linear function approximation, dynamical system identification and mechanical property prediction for hot rolled steels.

5.1. Comparative study of different cluster validity indices

To demonstrate the effectiveness of the proposed clustering validity measure, we compared the clustering performance associated with different validity criteria to the proposed clustering approach. To test the performance of different validity criteria, 400 data points consisting of four Gaussian clusters with 100 points per cluster, were generated as shown in Fig. 8. Five validity indexes: Partition Entropy *PE*, Partition Coefficient *PC*, Xie-Beni validity index $V_{\rm XB}$, Fukayama–Sugeno index $V_{\rm FS}$ and the proposed validity index $V_{\rm p}$ were used to partition the given data set. Table 1 displays the validation results of the five cluster validity indexes for $c = 2$–10 with different values of the fuzzy exponent m, which is considered to influence the validation. The highlighted cell values in the table refer to the optima detected for the corresponding indexes. It can be seen that all indexes point to the correct choice $c = 4$ when $m = 2$. As the value of m increased to 5, only the proposed index $V_{\rm p}$ selected the correct number of clusters while all others failed. When the value of m was decreased to 1.4, *PC* and $V_{\rm FS}$ were out of working order. With the value of $m = 1.2$, all validity measures lost their ability to validate the clusters. Fig. 9 shows the four Gaussian clusters contaminated by 200 randomly distributed noise data. The cluster validation results for the five validity indexes are listed in Table 2. It is seen that all validity indexes except the proposed validity index $V_{\rm p}$, failed to

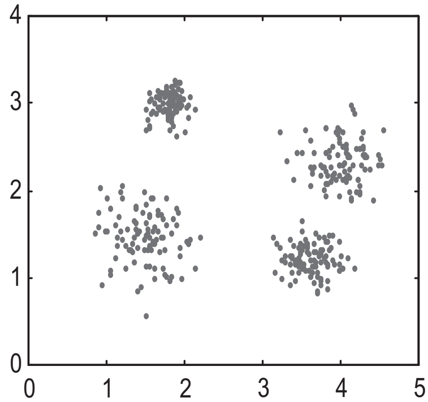

Fig. 8. Four Gaussian clusters with 100 points per cluster.

find the correct number $c=4$ when much noise is present in the data. It is clear that the proposed validity measure is more robust to random noise.

5.2. Function approximation

In this example, the fuzzy model was used to approximate the following two input-single output non-linear function, which is taken from [4,8,17].

$$y=(1+x_1^{-2}+x_2^{-1.5})^2. \tag{19}$$

The input data set consists of 100 randomly generated data points (x_1,x_2) in the range of $1\leqslant x_1\leqslant 5$ and $1\leqslant x_2\leqslant 5$. 50 data were used for training and the remaining 50 data for model testing. Using the proposed rule-base generation procedure, a 4-rule TSK fuzzy model was created and the corresponding parameters were trained using Eqs. (8)–(10). After parameter training, the MSE of the acquired model for testing was 0.0043. The membership functions of the input variables are displayed in Fig. 10(a). Obviously, there exist some highly overlapping fuzzy sets for both input variables. Using the approximate similarity analysis, the fuzzy sets with high similarity were merged and the simplified 4-rule model with 5 fuzzy terms is shown in Fig. 10(b). After parameter fine-tuning, the MSE of the final model for testing data was 0.0078. The model response surface is shown in Fig. 11. In addition to merging similar fuzzy sets, the 4-rule initial fuzzy model can be simplified to a 3-rule model by combining two similar rules with smaller threshold. After parameter re-tuning, the membership functions of the simplified 3-rule fuzzy model are shown in Fig. 10(c). From Fig. 10(a), we can also see that a fuzzy membership function of x_1 is always very small over its own universe of discourse. Thus, the initial 4-rule fuzzy model can be simplified further to a 2-rule fuzzy model (as show in Fig. 10(d)) by using both rule elimination and rule combination approaches presented in Section 4. The performance of the simplified models and the comparison with models developed in [4,8,17] are displayed in Table 3. It can be seen that the proposed fuzzy models not only improved its interpretability but also increased the model accuracy. It should be pointed out that the proposed fuzzy modelling approach can create fuzzy models with different degree of complexity.

Table 1
Comparison of clustering results for different partition validity (without noise data)

Indices	PE	PC	V_{XB}	V_{FS}	V_p
$m=1.2$					
$c=2$	**0.0065**	**0.9972**	0.2549	−0.6381	**0.9963**
$c=3$	0.0180	0.9918	0.3402	−0.9667	0.9937
$c=4$	0.0510	0.9717	1.0283	−1.1953	0.9782
$c=5$	0.0475	0.9726	0.1840	−1.5495	0.9793
$c=6$	0.0859	0.9508	0.3650	−1.5429	0.9648
$c=7$	0.1138	0.9340	0.3270	−1.5492	0.9527
$c=8$	0.1064	0.9402	0.2824	−1.5740	0.9591
$c=9$	0.1380	0.9209	0.2294	−1.5903	0.9443
$c=10$	0.1192	0.9342	**0.1589**	**−1.6040**	0.9543
$m=1.4$					
$c=2$	0.0331	**0.9880**	0.2663	−0.6130	0.9877
$c=3$	0.0602	0.9745	0.3208	−0.9795	0.9804
$c=4$	**0.0276**	0.9869	**0.0819**	−1.5845	**0.9909**
$c=5$	0.0741	0.9596	0.1788	−1.5847	0.9717
$c=6$	0.0910	0.9512	0.1788	**−1.5923**	0.9648
$c=7$	0.1566	0.9113	0.2967	−1.5617	0.9364
$c=8$	0.2046	0.8826	0.2409	−1.5617	0.9177
$c=9$	0.2131	0.8815	0.1537	−1.5813	0.9160
$c=10$	0.2444	0.8655	0.1720	−1.5699	0.9085
$m=2$					
$c=2$	0.3102	0.8304	0.2397	−0.4514	0.8059
$c=3$	0.3387	0.8328	0.1457	−1.0479	0.8595
$c=4$	**0.3021**	**0.8584**	**0.0702**	**−1.3953**	**0.8965**
$c=5$	0.4388	0.7892	0.1606	−1.2925	0.8438
$c=6$	0.5808	0.7109	0.1845	−1.1911	0.7813
$c=7$	0.6565	0.6861	0.1416	−1.1437	0.7672
$c=8$	0.7473	0.6421	0.1502	−1.0941	0.7329
$c=9$	0.7838	0.6313	0.1310	−1.0747	0.7273
$c=10$	0.8462	0.6110	0.1005	−1.0439	0.7151
$m=5$					
$c=2$	**0.6931**	**0.5000**	3.2997	0.1106	0.0045
$c=3$	0.9931	0.4083	0.0109	**−0.0698**	0.3129
$c=4$	1.2361	0.3364	0.0035	−0.0582	**0.3228**
$c=5$	1.5690	0.2164	0.0461	−0.0006	0.1095
$c=6$	1.7859	0.1687	0.0262	0.0012	0.0416
$c=7$	1.9340	0.1465	0.0061	0.0005	0.0488
$c=8$	2.0744	0.1263	0.0128	0.0004	0.0305
$c=9$	2.1913	0.1125	**0.0033**	0.0002	0.0309
$c=10$	2.2913	0.1024	0.0054	0.0001	0.0351

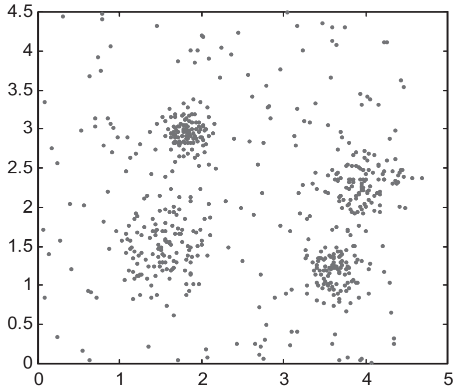

Fig. 9. Four Gaussian clusters contaminated with 200 noise points.

Table 2
Comparison of clustering results for different partition validity (with 200 noise data)

$m=2$	PE	PC	V_{XB}	V_{FS}	V_{p}
$c=2$	**0.3649**	**0.7789**	0.4071	−0.0207	0.7106
$c=3$	0.4946	0.7328	0.2965	−0.6238	0.7446
$c=4$	0.5340	0.7314	0.2031	−1.0028	**0.7713**
$c=5$	0.6726	0.6703	0.3471	−0.9558	0.7270
$c=6$	0.7145	0.6663	0.2595	**−1.0800**	0.7364
$c=7$	0.8067	0.6353	**0.1794**	−1.0444	0.7203
$c=8$	0.9224	0.5823	0.3007	−0.9625	0.6769
$c=9$	0.9551	0.5750	0.3377	−1.0344	0.6731
$c=10$	1.0358	0.5409	0.3010	−0.9696	0.6448

5.3. Dynamical system identification

The non-linear system studied in [16,22,25] is taken as the next example:

$$y(k) = g(y(k-1), y(k-2)) + u(k), \tag{20}$$

where

$$g(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)(y(k-1)-0.5)}{1+y^2(k-1)+y^2(k-2)}. \tag{21}$$

The system output depends on both its past values and the current input. The goal is to approximate the non-linear component $g(y(k-1), y(k-2))$ of the system with a fuzzy model. As in [16], 400 simulated data points were generated from the system model (21). Starting from the equilibrium state $(0,0)$, 200 samples of training data were obtained with a random input signal $u(k)$ uniformly distributed in $[-1.5, 1.5]$, followed by 200 samples of testing data obtained using a sinusoidal input signal $u(k) = \sin(2\pi k/25)$.

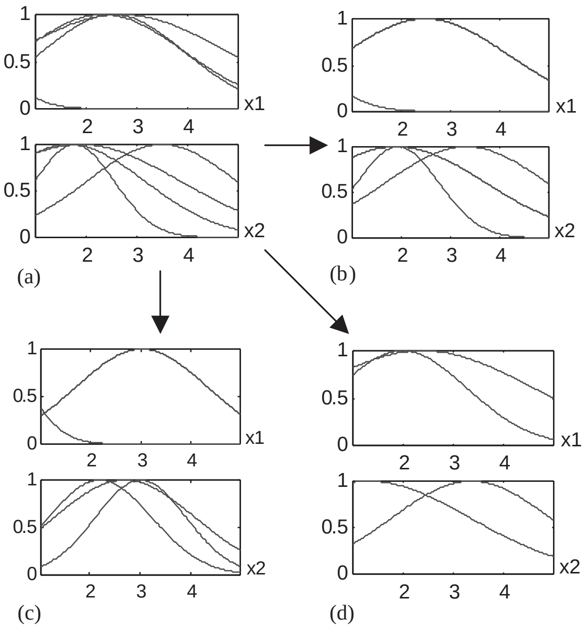

Fig. 10. Membership function distribution of Example B: (a) membership functions of the initial fuzzy model; (b) membership functions of the simplified 4-rule model; (c) membership functions of the simplified 3-rule model; (d) membership functions of the simplified 2-rule model.

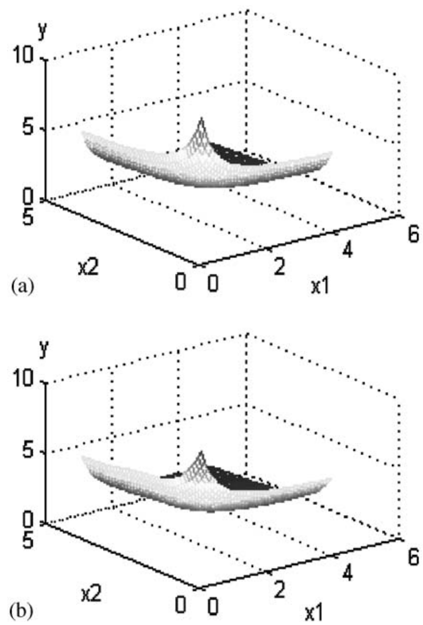

Fig. 11. Non-linear function approximation: (a) actual system output; (b) the output of the fuzzy model.

Table 3
Comparison of model performance for the non-linear function approximation

Ref.	No. of rules	No. of fuzzy sets	MSE
[17]	6 (initial)	12	0.318
	6 (final)	12	0.079
[4]	5 (initial)	10	0.314
	5 (final)	10	0.115
[8]	3 (final)	6	0.0197
This paper	4 (initial)	8	0.0043
	4 (simplified)	5	0.0078
	3 (simplified)	5	0.0191
	2 (simplified)	4	0.0756

In [25] several information criteria were applied to select rules from an initial model with 36 rules in order to obtain an optimal fuzzy model. The initial rule base was obtained by partitioning each of the two inputs $y(k-1)$ and $y(k-2)$ by six equidistant fuzzy sets. The rules were selected by their importance in the rule base determined by the rule reduction approach based on singular value decomposition. 1000 data were used for model training and 200 data used for testing. The obtained final optimized model consisted of 24 rules with 12 fuzzy sets.

In [22] an architecture of dynamic fuzzy neural networks (D-FNN) implementing TSK fuzzy systems was proposed. A hierarchical on-line self-organizing learning was used to generate rule-base. Fuzzy rules can be recruited or deleted dynamically according to their significance to the system's performance. The obtained final model consisted of 6 rules with 12 fuzzy sets. No fuzzy sets simplification was carried out in [22].

In [16] a Genetic Algorithm (GA)-based fuzzy modeling approach was proposed. An initial fuzzy model with 5 rules was generated by fuzzy clustering, and simplified by the set-theory based similarity measure. The final optimized model with 5 rules and 10 fuzzy sets was obtained by using genetic algorithms. A further simplified and optimized 4-rule TSK model also showed good performance.

Using the proposed approach in this paper, a 5-rule fuzzy model with 5 fuzzy sets for each input variable was generated. Fig. 12(a) illustrates the distribution of the membership functions on the $y(k-1)$ and $y(k-2)$ dimensions. It is easy to see that some membership functions have high similarity degree and can be merged together. After rule-base simplification and parameter fine-tuning, a simplified model with only 6 membership functions (3 fuzzy sets for $y(k-1)$ and 3 fuzzy sets for $y(k-2)$, respectively) was obtained, as shown in Fig. 12(b). In research of further simplification of the model, we reduced the merging threshold and generated a reduced 4-rule model with 5 fuzzy sets. After fine-training, the model accuracy kept in an acceptable level, MSE $= 1.1e^{-3}$ and $3.7e^{-4}$ for training and testing, respectively. We compared the performances of the proposed fuzzy model to those in the literatures. The comparative results are listed in Table 4. It can be seen that compared to the fuzzy models in the literatures, the proposed models keep good balance between numerical accuracy and model simplicity.

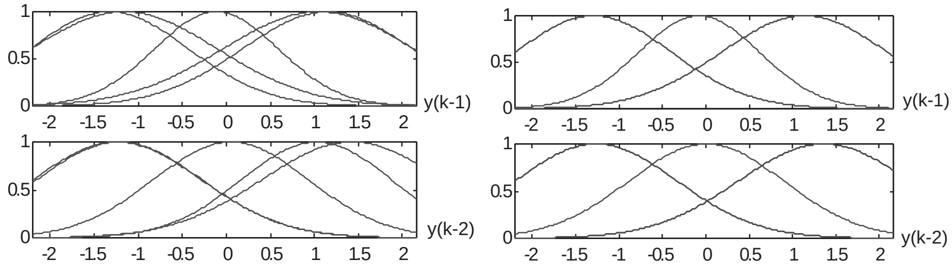

Fig. 12. The distribution of membership functions on the $y(k-1)$ and $y(k-2)$: (a) original fuzzy model; (b) simplified fuzzy model.

Table 4
Comparative results of different fuzzy models for dynamical system modeling

Ref.	No. of rules	No. of fuzzy sets	MSE train	MSE test
[25]	36 (initial)	12	$1.9e^{-6}$	$2.9e^{-3}$
	24 (final)	12	$2.0e^{-6}$	$6.4e^{-4}$
[22]	6 (final)	12	—[a]	$8.0e^{-4}$
[16]	5 (initial)	10	$5.8e^{-3}$	$2.5e^{-3}$
	5 (final)	8	$7.5e^{-4}$	$3.5e^{-4}$
	4 (final)	4	$1.2e^{-3}$	$4.7e^{-4}$
This paper	5 (initial)	10	$1.9e^{-4}$	$3.8e^{-4}$
	5 (final)	6	$1.0e^{-4}$	$3.2e^{-4}$
	4 (final)	5	$1.1e^{-3}$	$3.7e^{-4}$

[a]The result is not listed in the original paper.

5.4. Mechanical property prediction for hot rolled steels

The problem in modeling the properties of hot-rolled metal materials can be broadly stated as: given a certain material which undergoes a specified set of manufacturing processes, what are the final properties of this material? Typical final mechanical properties in which we are interested are tensile strength, yield stress, elongation, etc. By using the proposed fuzzy modeling approach, we have developed composition-microstructure-property models for a wide range of hot-rolled steels. 601 industrial data from carbon-manganese steels and niobium microalloyed steels have been used to train and test the fuzzy model, which relates the chemical compositions and microstructure to the mechanical properties. Six inputs (carbon, silicon, manganese, nitrogen and niobium contents and ferrite grain size $D^{-1/2}$) were selected from the 14 possible input variables. To determine the number of fuzzy rules, the proposed fuzzy clustering method was used to find out the data structure and the optimal number of clusters. Different cluster validity measures have been used on this data set and the corresponding clustering results are listed in Table 5. It can be seen that *PE* and *PC*

Table 5
Comparison of clustering performance for different validity criteria (601 industrial data)

$m=2$	PE	PC	V_{XB}	V_{FS}	V_p
$c=2$	**0.2702**	**0.8374**	8.9513	−980.6	0.7120
$c=3$	0.4119	0.7704	6.6778	−1804.8	0.7447
$c=4$	0.4607	0.7593	4.6909	−2088.2	0.7803
$c=5$	0.5468	0.7263	3.6978	−2136.2	0.7650
$\mathbf{c=6}$	0.5516	0.7304	3.3155	−2223.8	**0.7890**
$c=7$	0.6048	0.7118	3.0335	−2229.7	0.7826
$c=8$	0.6150	0.7099	2.6182	−2209.7	0.7842
$c=9$	0.6399	0.7085	**2.2865**	−2212.0	0.7817
$c=10$	0.6526	0.6996	2.4062	**−2235.6**	0.7820

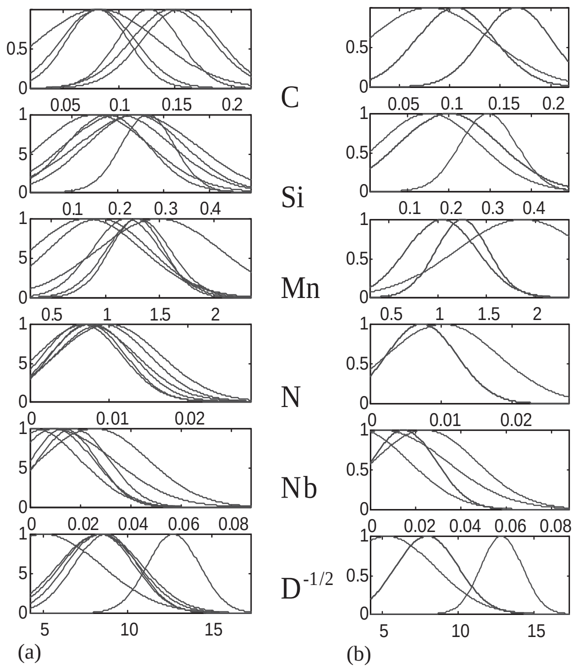

Fig. 13. Membership functions of the property prediction model.

monotonically tend to the minimum number of clusters, i.e. $c_{\min}=2$, while V_{XB} and V_{FS} are close to the maximum number $c_{\max}$. Only the proposed index V_p selected $c=6$, which is consistent with expert recommendation and experimental results. It is shown that the V_p is more reliable on high-dimensional data.

After rule-base generation and parameter learning, a 6-rule fuzzy model of the Mamdani type was obtained. The distribution of the membership functions for each input variable is represented in Fig. 13(a). Using this model, we obtained the *standard deviation* (*SD*) of the prediction error:

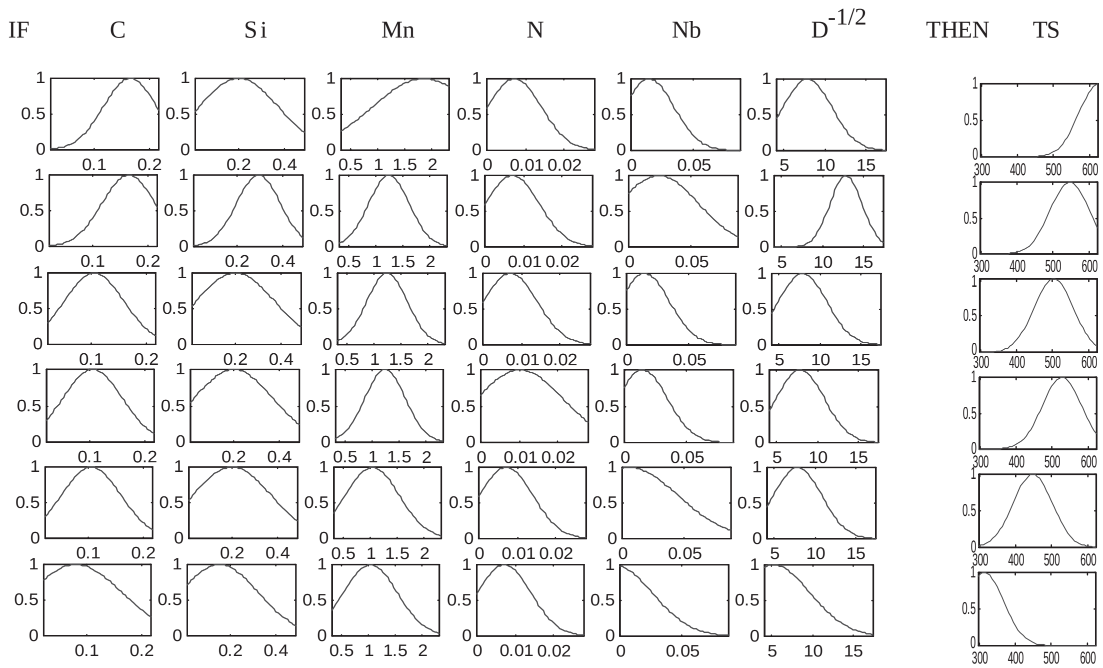

Fig. 14. Final fuzzy model for Nb-alloy steels.

$SD = 15.16$ and 19.74 for training (301 data) and testing (300 data), respectively. Again, we can see some similar fuzzy sets existing in the generated model as shown in Fig. 13(a). Obviously, this model can be further simplified to a more transparent model displayed in Fig. 13(b) via the approximate similarity measures proposed in the previous sections. After fine-tuning, the final fuzzy model with $SD = 16.94$ and 21.65 (for training and model testing, respectively) was obtained as shown in Fig. 14. From this fuzzy model, we can use the linguistic hedges approach [5] to derive the corresponding interpretable linguistic model as follows:

R_1: If C is *quite large* and Si is *more or less medium* and Mn is *more or less large* and N is *slightly small* and Nb is *slightly small* and $D^{-1/2}$ is *slightly small*, Then TS is *large*.

R_2: If C is *quite large* and Si is *medium* and Mn is *medium* and N is *slightly small* and Nb is *more or less small* and $D^{-1/2}$ is *slightly large*, Then TS is *quite large*.

R_3: If C is *more or less medium* and Si is *more or less medium* and Mn is *medium* and N is *slightly small* and Nb is *small* and $D^{-1/2}$ is *slightly small*, Then TS is *medium*.

R_4: If C is *more or less medium* and Si is *more or less medium* and Mn is *medium* and N is *more or less medium* and Nb is *small* and $D^{-1/2}$ is *slightly small*, Then TS is *quite large*.

R_5: If C is *more or less medium* and Si is *more or less medium* and Mn is *medium* and N is *slightly small* and Nb is *more or less small* and $D^{-1/2}$ is *slightly small*, Then TS is *medium*.

R_6: If C is *more or less small* and Si is *more or less small* and Mn is *medium* and N is *slightly small* and Nb is *more or less small* and $D^{-1/2}$ is *more or less small*, Then TS *is small*.

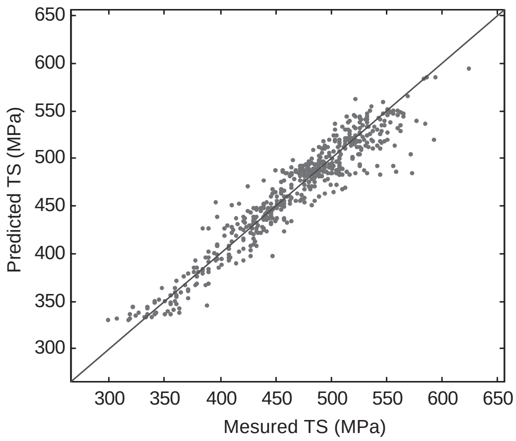

Fig. 15. Model predicted TS versus measured TS.

It is worth noting that the obtained fuzzy rule-based model reveals relationships between composition-microstructure and tensile strength, which are consistent with metallurgical knowledge. The performances of the final simplified model is shown in Fig. 15. Thus it has been shown that the simplified fuzzy model has a much simpler structure and better interpretability with only slight loss of accuracy.

6. Conclusions

In this paper, a data-driven fuzzy modeling and simplification approach is proposed. Using this approach, an interpretable fuzzy rule-based model can be generated and optimized automatically from the training data. Due to its multi-paradigm nature, including fuzzy clustering, partition validation, approximate similarity analysis and parameter learning, the obtained model not only provides an interpretable and simple model structure but also maintains good model accuracy. Thus we obtain a method for finding a good balance between model accuracy and model transparency. The simplified rule base is computationally efficient and linguistically tractable, from which it may also reveal a useful qualitative description of the system that generated the data. Such a description can be examined and possibly combined with the knowledge of experts, helping to understand the system and validate the model at the same time. Experimental validation shows that the produced rule-based fuzzy models have satisfactory prediction accuracy and good interpretation features. Clearly, the proposed fuzzy modeling approach provides a simple and effective framework for system identification and prediction. Further improvement in the model optimization and incorporation of a priori physically-based linguistic information into the modeling procedure would be beneficial.

References

[1] J.C. Bezdek, Cluster validity with fuzzy sets, J. Cybernet. 3 (3) (1974) 58–72.
[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum, New York, 1981.

[3] C.T. Chao, Y.J. Chen, C.C. Teng, Simplification of fuzzy neural systems using similarity analysis, IEEE SMC-B 26 (2) (1996) 344–354.

[4] M. Delgado, A.F. Gomez-Skarmeta, F. Martin, A fuzzy clustering-based rapid prototyping for fuzzy rule-based modelling, IEEE Trans. Fuzzy Systems 5 (2) (1997) 223–233.

[5] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy c-means method, Proc. 5th Fuzzy Syst. Symp., Japan, 1989, 247–250.

[6] J.R. Jang, ANFIS: adaptive network-based fuzzy inference system, IEEE Trans. Syst. Man Cybernet. 23 (4) (1993) 665–684.

[7] Y. Jin, Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement, IEEE Trans. Fuzzy Systems 8 (2) (2000) 212–221.

[8] E. Kim, M. Park, S. Ji, M. Park, A new approach of fuzzy modelling, IEEE Trans. Fuzzy Systems 5 (3) (1997) 328–337.

[9] C.T. Lin C.S.G. Lee, Reinforcement structure/parameter learning for neural-network based fuzzy logic control systems, IEEE Trans. Fuzzy Systems 2 (1) (1994) 46–63.

[10] S. Marsili-Libelli, A. Muller, Adaptive fuzzy pattern recognition in the anaerobic digestion process, Pattern Recognition Lett. 17 (6) (1996) 651–659.

[11] V. de Oliveira, Semantic constraints for membership function optimization, IEEE Trans. SMC-A 29 (1) (1999) 128–138.

[12] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy c-means model, IEEE Trans. Fuzzy Systems 3 (3) (1995) 370–379.

[13] S.J. Qin, G. Borders, A multiregion fuzzy logic controller for nonlinear process control, IEEE Trans. Fuzzy Systems 2 (1) (1994) 74–81.

[14] M. Setnes, R. Babuska, U. Kaymak, H.R.N. Lemke, Similarity measures in fuzzy rule base simplification, IEEE Trans. SMC-B 28 (3) (1998) 376–386.

[15] M. Setnes, R. Babuska, H. Verbruggen, Rule-based modelling: precision and transparency, IEEE Trans. SMC-C 28 (1) (1998) 165–170.

[16] M. Setnes, H. Roubos, GA-fuzzy modeling and classification: complexity and performance, IEEE Trans. Fuzzy Systems 8 (5) (2000) 509–522.

[17] M. Sugeno, T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling, IEEE Trans. Fuzzy Systems 1 (1) (1993) 7–31.

[18] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Trans. Syst. Man Cybernet. 15 (1) (1985) 116–132.

[19] E. Vandewalle, Constructing fuzzy models with linguistic integrity from numerical data—AFRELI algorithm, IEEE Trans. Fuzzy Systems 8 (5) (2000) 591–600.

[20] L.X. Wang, Adaptive Fuzzy Systems and Control, Prentice-Hall, Englewood Cliffs, NJ, 1994.

[21] Li.X. Wang, Modelling and control of hierarchical systems with fuzzy systems, Automatica 33 (6) (1997) 1041–1053.

[22] S. Wu, M.J. Er, Dynamic fuzzy neural networks—a novel approach to function approximation, IEEE Trans. SMC-B 30 (2) (2000) 358–364.

[23] X.L. Xie, G.A. Beni, Validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (8) (1991) 841–846.

[24] J. Yen, L. Wang, Simplifying fuzzy rule-base models using orthogonal transformation methods, IEEE Trans. SMC-B 29 (1) (1999) 13–24.

[25] J. Yen, L. Wang, Constructing optimal fuzzy models using statistical information criteria, J. Intell. Fuzzy Systems 7 (2) (1999) 185–202.