

## HIERARCHICAL CLUSTERING FOR FUZZY MODELING OF MATERIALS PROPERTY PREDICTION

Min-You Chen and D. A. Linkens

*Dept. of Automatic Control and Systems Engineering  
University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK.  
Email: [Minyou.chen@shef.ac.uk](mailto:Minyou.chen@shef.ac.uk) [d.linkens@shef.ac.uk](mailto:d.linkens@shef.ac.uk)*

**ABSTRACT:** A simple and effective fuzzy clustering approach is presented for fuzzy modeling from industrial data. In this approach, fuzzy clustering is implemented in two phases: data compression by a self-organizing network, and fuzzy partitioning via fuzzy  $c$ -means clustering associated with a proposed cluster validity measure. The approach is used to extract fuzzy models from data and find out the optimal number of fuzzy rules. The simulation results show that the proposed approach has good clustering performance with noise-contaminated data and high-dimensional industrial data. *Copyright © 2002 IFAC*

**Keywords:** Fuzzy clustering, Partition validity, Fuzzy modeling, Property prediction.

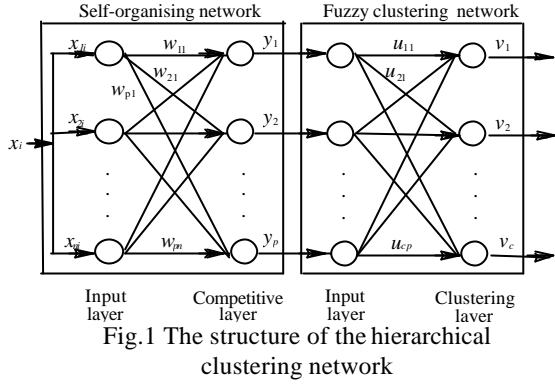
### 1. INTRODUCTION

Cluster analysis has been playing an important role in solving problems in many engineering applications, such as data mining, pattern recognition, image processing and extraction of fuzzy rules from data. The basic objective of cluster analysis is to partition the  $N$  unlabeled data points into  $c$  clusters in an optimal fashion. In various developed clustering algorithms, the fuzzy  $c$ -means (FCM) algorithm is the most commonly used algorithm due to its efficacy and simplicity. However, FCM suffers from two challenging problems: unknown number of clusters and noise contaminated data. The first is that the number  $c$  of clusters must be pre-defined and the resulting structure for the specified number of clusters is assumed to be the best. This is seldom the case in practice. Thus, the difficult problem encountered is the cluster validity which is required to evaluate the quality of the  $c$ -partitions resulting from the algorithms. The second is that the FCM algorithm is sensitive to noise in the data. When noise is present in the data set, the FCM algorithm can give distorted results or even fail completely. To improve the performance of clustering, various clustering algorithms have been proposed for dealing with either unknown number of clusters or noise data problems. However, most of the clustering algorithms focusing on improving robustness or extending the function of FCM (Kirshnapuram and Keller 1993, Dave 1993, Pedrycz 1996, Nasraoui and Kirshnapuram 1996), are computationally intensive. Little attention has been paid to improving the

computational efficiency of the algorithms. In this paper, a *hierarchical fuzzy clustering* (HFC) approach associated with a simple and effective cluster validity criterion is presented. The clustering procedure is implemented by two hierarchically connected networks: a self-organizing clustering network and a FCM-based fuzzy clustering network. The structure of the hierarchical clustering network proposed in (Linkens & Chen 1998) and corresponding clustering algorithms associated with a new cluster validity criterion will be presented in Section 2. In Section 3, the comparison among different validity indices is given to demonstrate the effectiveness of the proposed validity measure. Application to fuzzy modeling for material property prediction are provided in Section 4. Finally, concluding remarks are given in Section 5.

### 2. HIERARCHICAL FUZZY CLUSTERING

Given a set of unlabeled data  $X = \{x_1, x_2, \dots, x_n\}$ , the objective of fuzzy clustering is to find the best partition of  $n$  entities into  $c$  clusters. In contrast to most existing clustering algorithms which generate  $c$  clusters directly from raw data, the proposed approach consists of two hierarchical stages: generating a relatively small number of  $p \ll n$  initial sub-clusters via a competitive network, and then partitioning  $p$  sub-clusters into  $c$  clusters by a fuzzy clustering network under the proposed partition validity criterion. The structure of the hierarchical clustering network is depicted in Fig. 1. The procedure of the hierarchical clustering is discussed in the following subsections.



## 2.1 Data compression using self-organizing network

The self-organizing network (Kohonen 1995) is introduced to form the first hierarchy so as to produce the sub-clusters based on the given data set. The purpose of this phase is to classify the given training data into a small number, say  $p \ll n$ , clusters using competitive learning. It is known that FCM algorithms deal with  $p \times n$  partition matrices based on raw data, where  $p$  is the number of clusters and  $n$  is the number of input vectors. In practical problems  $n$  may be  $10^3$  or larger, so that computation of partition matrices becomes overwhelming. The first hierarchy is used as a pre-processor to perform fuzzy classification with the dual objectives of providing more representative information and reducing the total number of training instances for the subsequent FCM algorithms. The number of clusters is determined by the classification stage. Since both the competitive network and FCM algorithm lack good noise rejection ability, a data density based noise-removing rule is introduced in the algorithms. The modified self-organizing algorithm is presented as follows:

### Step 1. Network initialization

Given unlabelled data set  $X = \{x_1, x_2, \dots, x_n\} \subset R^s$ . Process the first input pattern  $x_1 = (x_{11}, x_{21}, \dots, x_{s1})$ , set the iteration number  $l=1$ . Let the first weight vector be  $W_1^1 = x_1$ , i.e.  $w_{i1}^1 = x_{i1}$ ,  $i=1, 2, \dots, n$ . Specify the valid radius  $d$  for all neurons. Set the number of neurons  $N_t=1$ , and the activation number of node 1  $N_{S1}=1$ ;

**Step 2.** For the  $l$ th input, Find the node  $J$  which has the minimum distance to the current input pattern  $x_l$  by  $D(W_j^l, x_l) = \|W_j^l - x_l\| = \min_i \|W_i^l - x_l\|$ ,  $i=1, 2, \dots, N_t$ .

The distance is defined as

$$\|w_i^l - x_l\| = (w_i^l - x_l)(w_i^l - x_l)^T$$

**Step 3.** Determine the winner using the following rule:

$$\text{If } \begin{cases} D(w_j^l, x_l) \leq d, & \rightarrow J \text{ is the winner} \\ D(w_j^l, x_l) > d, & \rightarrow \text{create a new node} \end{cases}$$

If  $J$  is the winner, modify the weight vector of unit  $J$  to:

$W_j^l = W_j^{l-1} + \alpha \|x_l - W_j^l\|$ , where  $\alpha$  is the learning rate which is determined by

$\alpha = \alpha_0 / (N_{SJ} + 1)$ , where  $\alpha_0 \in [0, 1]$  is the initial rate;  $N_{SJ} = N_{SJ} + 1$ ;  $l = l + 1$ .

If creating a new unit, then the weight vector is given by  $W_n^l = x_l$ ,  $N_t = N_t + 1$ ;

If  $l < n$ , go to step 2, otherwise set  $p = N_t$  and go to next step.

### Step 4. Output selection

Set  $N_m = \max\{N_{S1}, N_{S2}, \dots, N_{Sp}\}$  in the competitive layer, remove the nodes whose activation number  $N_{Sj} < rN_m$ , where  $r \in (0, 1)$  is a constant factor,  $j \in \{1, 2, \dots, p\}$ . Reset  $p = p - N_r$ , where  $N_r$  denotes the number of removed nodes. The activation value of an output node is defined as:  $y_j = W_j$ ;  $j=1, 2, \dots, p$ , where  $W_j = (w_{j1}, w_{j2}, \dots, w_{jn})$  represents the prototype of the  $j$ th fuzzy cluster in input-output space.

Obviously, the clusters with low point density (i.e. with low values of  $N_s$ ) will be removed through step 4. This means that noise-corrupted points will be rejected in forming the sub-clusters because of their random distribution. The created sub-clusters will represent the data structure well.

Once the unsupervised learning is completed, a collection of  $p$  fuzzy clusters  $Y = \{y_1, y_2, \dots, y_p\}$  represented by the  $p$  nodes in the competitive layer is produced.

Based on the  $p$  sub-clusters, the second hierarchy, i.e. the fuzzy clustering, involves two tasks: 1) to decide the optimal number  $c$  of clusters; 2) to find out the best partition of  $c$  clusters. The former is related to the cluster validity problem and the latter refers to the fuzzy clustering problem. We will discuss these problems in the following two subsections.

## 2.2 Cluster Validity

The problem of partition validation is to find the best number of clusters under certain cluster validity criterion. It is noted that in the case of rule extraction from data, an effective partition in input-output space can lead to reducing the number of rules and thus improving the computational efficiency and interpretability of the fuzzy model. Numerous clustering algorithms have been developed. The most widely used algorithm is the fuzzy  $c$ -means (FCM) due to its efficacy and simplicity. However, the number  $c$  of clusters is required to be pre-determined. The FCM algorithm partitions a collection of  $n$  data points ( $X = \{x_1, x_2, \dots, x_n\}$ ) into  $c$  fuzzy clusters such that the following objective function is minimized.

$$J_m = \sum_{k=1}^n \sum_{i=1}^c m_k^m(x) \|x_k - v_i\|^2, \quad 1 < m < \infty \quad (1)$$

where  $m$  is a fuzzy coefficient,  $v_i$  is the prototype of the  $i$ th cluster generated by fuzzy clustering,  $u_{ik}$  is the membership degree of the  $k$ th data belonging to the  $i$ th cluster represented by  $v_i$ ,  $u_{ik} \in U$ ,  $U$  is a  $c \times n$  fuzzy partition matrix which satisfies the constraints:

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

$$\sum_{k=1}^n u_{ik} = 1 \quad \text{for } k=1, 2, \dots, n$$

Since  $J_m$  monotonically decreases with  $c$ , an efficient criterion for evaluating the performance is required. Many cluster-validity criteria have been proposed to measure the effectiveness of the clustering. The first fuzzy cluster-validity criteria associated with the FCM introduced by Bezdek are the *partition coefficient* (PC) and the *partition entropy* (PE) (Bezdek 1974, 1981). Their main advantage is their simplicity but the main disadvantage is their monotonic evolution tendency with respect to  $c$ . Fukayama & Sugeno (1993) and Xie & Beni (1992) have 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 would lose their ability to validate partitions from FCM for large  $m$ . In this study, a simple and effective alternative fuzzy partition measure is proposed as a cluster validity criterion associated with the FCM algorithm, which is defined as follows:

$$V_p(U) = \frac{1}{n} \sum_{k=1}^n \max_i(u_{ik}) - \frac{1}{K} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[ \frac{1}{n} \sum_{k=1}^n \min(u_{ik}, u_{jk}) \right]$$

$$\text{where } K = \sum_{i=1}^{c-1} i \quad (2)$$

It can be seen that the cluster validity measure  $V_p$  is composed of two items. The first item reflects the compactness within a cluster. The closer the  $k$ th pattern  $x_k$  is to a fuzzy cluster centre, the closer the maximum membership degrees  $\max(u_{ik})$  is to the value 1. Hence, the fuzzy set  $\max(u_{ik})$  is considered as a good indicator of the clustering quality for each pattern  $x_k$ . This quality indicates how closely the objects are assigned to the fuzzy cluster centres. Thus, a large value of the first item indicates the data patterns are well classified. On the other hand, the second item indicates that the separation between clusters. Here, the intersection of two fuzzy sets is used to evaluate a fuzzy separation between clusters  $v_i$  and  $v_j$ . In fact, if  $x_k$  is close to the fuzzy cluster centre of  $v_i$ ,  $\min(u_{ik}, u_{jk})$  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  $1/c$ ,  $x_k$  belongs to all clusters with equal value and the fuzziest separation is obtained. The new validity  $V_p$  criterion combines information about fuzzy compactness and separation. It tends to indicate a good cohesion within clusters and a small overlap between pairs of clusters. It is worth noting that the proposed index  $V_p$  is computationally simple as it involves only maximum and minimum values of the partition matrix  $U$ . In contrast to the indexes PC and PE,  $V_p$  is independent of  $c$ , which can be seen in its limiting behavior.

$$\lim_{m \rightarrow 1} \{V_p(U)\} = \lim_{m \rightarrow 1} \left\{ \frac{1}{n} \sum_{k=1}^n \max_i(u_{ik}) \right\}$$

$$- \lim_{m \rightarrow 1} \left\{ \frac{1}{K} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[ \frac{1}{n} \sum_{k=1}^n \min(u_{ik}, u_{jk}) \right] \right\} = 1$$

$$\lim_{m \rightarrow \infty} \{V_p(U)\} = \lim_{m \rightarrow \infty} \left\{ \frac{1}{n} \sum_{k=1}^n \max_i(u_{ik}) \right\}$$

$$- \lim_{m \rightarrow \infty} \left\{ \frac{1}{K} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[ \frac{1}{n} \sum_{k=1}^n \min(u_{ik}, u_{jk}) \right] \right\} = 0$$

It is clear that the  $V_p$  overcomes the main shortcoming of PC and PE, which have monotonic evolution tendency with respect to  $c$ . However, as the value of  $V_p$  converge to a constant when the fuzzy exponent  $m$  is very close to one or very large (towards infinity), it would lose its ability to discriminate between various values of  $c$ . According to experimental results, the proposed validity index  $V_p$  works very well in the range of  $m \in [1.5, 5]$ . Fortunately, the values of  $m < 5$  are very usual in practice, and  $m=2$  is so far the most common choice.

### 2.3 Fuzzy clustering procedure

FCM clustering is performed via a clustering network as shown in Fig. 1. It is noted that the training data of the FCM clustering are prototypes of the  $p$  sub-clusters generated by the self-organizing network instead of the raw data set if the number of raw data is very large. The clustering network based on a FCM algorithm attempts to classify the given set of data points 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 (2) is carried out through an iterative optimization of  $J_m$  according to the following steps:

*Step 1.* Choose the maximum cluster number  $c_{max}$ , iteration limit  $T$ , weighting exponent  $m$ , and termination criterion  $\epsilon > 0$ .

*Step 2.* With  $c=2, 3, \dots, c_{max}$ , initialize the position of cluster centres:  $V_0 = (v_{10}, v_{20}, \dots, v_{c0})$ ;

*Step 3.* With the iteration number  $t=1, 2, \dots, T$ ;

$$\text{calculate; } u_{ik,t} = 1 / \sqrt{\sum_{j=1}^c (d_{ik} / d_{jk})^{2/(m-1)}} \quad (3)$$

where  $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|$ ,  $i=1, 2, \dots, c$ ;  $k=1, 2, \dots, p$ ;

$$\text{Calculate } \mathbf{v}_{i,t} = \sum_{k=1}^n (u_{ik,t})^m \mathbf{x}_k / \sum_{k=1}^n (u_{ik,t})^m \quad (4)$$

If  $\|\mathbf{V}_t - \mathbf{V}_{t-1}\| < \varepsilon$ , go to next step, otherwise repeat step 3.  
*Step 4.* Calculate  $V_p(c)$  by (2); if  $c < c_{max}$ , repeat from Step 2. Otherwise, stop the program and set the optimal cluster number  $c = c_m$ , where  $c_m$  meets the following condition:

$$V_p(c_m) = \max\{V_p(c)\}, c=2, 3, \dots, c_{max};$$

Based on cluster validation, both the number of rules and the prototypes of the clusters  $\mathbf{v}_j = (v_{j1}, v_{j2}, \dots, v_{js}, v_{j,s+1})$  can be obtained, where  $j=1, 2, \dots, c$ .

It is easy to see that the computational complexity of the FCM algorithm in the second level of the hierarchy, in which the main computations are related to calculating  $u_{ik}$  and  $v_i$  represented by Eq. (3) and (4), is higher than that of the self-organizing network in the first level of the hierarchy, in which the most complex computation is just related to distance calculation. Suppose  $t^*$  is the number of iterations for termination. In FCM algorithms, its complexity corresponding to  $u_{ik}$  computation is  $O(t^*n)$ , where  $n$  is the number of data points, while in HFC algorithms the computational complexity related to  $u_{ik}$  is  $O(t^*p)$  because all computations are based on the  $p$  representative prototype points instead of  $n$  data points, (heuristically,  $\sqrt{n} \leq p \ll n$ ). Thus, the computational cost in the HFC algorithms is decreased drastically.

### 3. COMPARISON OF DIFFERENT CLUSTER VALIDITY INDICES

To demonstrate the effectiveness of the proposed clustering approach, 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.2(a). Five validity indexes: Partition Entropy  $PE$ , Partition Coefficient  $PC$ , Xie-Beni validity index  $V_{XB}$ , Fukayama-Sugeno index  $V_{FS}$  and the proposed validity index  $V_p$  were used to partition the given data set. Table 1 displays the validation results of the five cluster validity indexes for  $c=2$  to 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 by 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  $V_p$

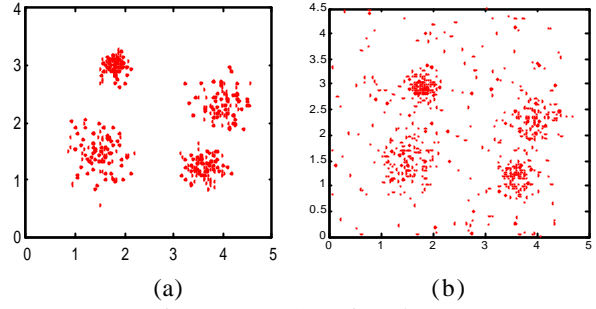


Fig. 2. Four Gaussian clusters.

(a) without noise data; (b) with 200 noise data

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

Index	$PE$	$PC$	$V_{XB}$	$V_{FS}$	$V_p$
$m=1.4$					
C=2	0.033	<b>0.988</b>	0.266	-0.613	0.988
C=3	0.060	0.975	0.321	-0.979	0.980
<b>C=4</b>	<b>0.028</b>	0.987	<b>0.082</b>	-1.585	<b>0.991</b>
C=5	0.074	0.960	0.179	-1.585	0.972
C=6	0.091	0.951	0.179	<b>-1.592</b>	0.965
C=7	0.157	0.911	0.297	-1.562	0.936
C=8	0.205	0.883	0.241	-1.562	0.918
C=9	0.213	0.882	0.154	-1.581	0.916
C=10	0.244	0.865	0.172	-1.570	0.909
$m=2$					
C=2	0.310	0.830	0.240	-0.451	0.806
C=3	0.339	0.833	0.146	-1.048	0.860
<b>C=4</b>	<b>0.302</b>	<b>0.858</b>	<b>0.070</b>	<b>-1.395</b>	<b>0.897</b>
C=5	0.439	0.789	0.161	-1.293	0.844
C=6	0.581	0.711	0.185	-1.191	0.781
C=7	0.657	0.686	0.142	-1.144	0.767
C=8	0.747	0.642	0.150	-1.094	0.733
C=9	0.784	0.631	0.131	-1.075	0.727
C=10	0.846	0.611	0.101	-1.044	0.715
$m=5$					
C=2	<b>0.693</b>	<b>0.500</b>	3.300	0.111	0.005
C=3	0.993	0.408	0.011	<b>-0.070</b>	0.313
<b>C=4</b>	1.236	0.336	0.004	-0.058	<b>0.323</b>
C=5	1.569	0.216	0.046	-0.001	0.110
C=6	1.786	0.169	0.026	0.001	0.042
C=7	1.934	0.147	0.006	0.0005	0.049
C=8	2.074	0.126	0.013	0.0004	0.031
C=9	2.191	0.113	<b>0.003</b>	0.0002	0.031
C=10	2.291	0.102	0.005	0.0001	0.035

selected the correct number of clusters while all others failed. When the value of  $m$  was decreased to 1.4,  $PC$  and  $V_{FS}$  were out of working order. Fig.2(b) 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  $V_p$ , failed to 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.

Table 2. Comparison of Clustering performance for different partition validity (with 200 noise data)

$m=2$	$PE$	$PC$	$V_{XB}$	$V_{FS}$	$V_p$
C=2	<b>0.365</b>	<b>0.779</b>	0.407	-0.021	0.711
C=3	0.495	0.733	0.296	-0.624	0.745
<b>C=4</b>	0.534	0.731	0.203	-1.003	<b>0.771</b>
C=5	0.673	0.670	0.347	-0.956	0.727
C=6	0.715	0.666	0.259	<b>-1.080</b>	0.736
C=7	0.807	0.635	<b>0.179</b>	-1.044	0.720
C=8	0.922	0.582	0.301	-0.963	0.677
C=9	0.955	0.575	0.338	-1.034	0.673
C=10	1.036	0.541	0.301	-0.970	0.645

#### 4. FUZZY MODELING FOR MATERIAL PROPERTY PREDICTION

The problem in modeling the properties of hot-rolled alloy steels 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 strength and toughness. In this study, the hierarchical clustering approach was applied to extract fuzzy rule-based models from industrial data for tensile strength prediction of heat-treated structural steels. Based on a collection of  $N$  input/output data points  $\mathbf{P}(x, y)$ , a number of  $p$  clusters can be generated using the proposed clustering approach. Each cluster center  $\mathbf{c}_i$  ( $i=1,2,\dots,p$ ) is in essence a prototypical data point that exemplifies a characteristic input/output behavior of the system we wish to model. Hence each cluster center can be used as the basis of a rule that describes the system behavior. Each vector  $\mathbf{c}_i$  can be decomposed into two component vectors:  $\mathbf{c}_i=(x_i^* \quad y_i^*)$ , where  $x_i^*=(x_{i1}^*, x_{i2}^*, \dots, x_{in}^*)$ . Thus, each cluster center can be viewed as a fuzzy rule that describes the system's local behavior. Intuitively, cluster center  $\mathbf{c}_i$  represents the rule "IF input is around  $x_i^*$  THEN output is around  $y_i^*$ ". Hence, the fuzzy rule-base consisting of  $p$  rules is created by the fuzzy clustering. The deviation parameter of cluster  $i$ ,  $\mathbf{s}_i=\mathbf{s}_i$ , can be decided by using the average distance to the nearest  $m$ -cluster centers:

$$\mathbf{s}_i = \left( \sum_{j=1}^m \|\mathbf{c}_i - \mathbf{c}_j\| / m \right)^{1/2} \quad (9)$$

where  $\mathbf{c}_j$  is the centre of the  $j$ th cluster near to the cluster  $i$ . The obtained  $p$  prototypes can be used to construct the parameters of the fuzzy rule-base.

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

$R_j$ : If  $x_1$  is  $A_{j1}$  and  $x_2$  is  $A_{j2}$ ... and  $x_n$  is  $A_{jn}$  Then  $y$  is  $z_j$  where  $R_j$  denotes the  $j$ th rule,  $j=1,2,\dots, p$ ;  $A_{ji}$  is the fuzzy set defined by the Gaussian membership function;  $z_j$  is the  $j$ th rule output with respect to the fuzzy model. All model parameters can be optimized by gradient-descent based learning algorithm.

Using the clustering-based fuzzy modeling approach, we have developed Tensile Strength (TS) prediction models for heat-treated structural steels. 3804 industrial data from carbon-manganese-nickel alloyed steels have been used to train and test the fuzzy model, which relates the chemical compositions and process conditions to the mechanical properties. Nine inputs (C, Si, Mn, Ni, Cr, Mo, V, Gauge and Tempering Temperature) were selected from the 19 possible input variables. To determine the number of fuzzy rules, the proposed fuzzy clustering method was applied to find out the data structure and the optimal number of clusters. Different cluster validity measures have been used on this data set for comparison and the corresponding clustering results are listed in Table 3. It can be seen that  $PE$  and  $PC$  tend to the maximum and minimum number of clusters, i.e.  $c=2$  and  $c=10$ , while  $V_{FS}$  chooses  $c=9$ . Only  $V_{XB}$  and the proposed index  $V_p$  select  $c=5$ , which is consistent with expert recommendation and experimental results. It is shown that the  $V_p$  is quite reliable on high dimensional data.

Table 3. Clustering results for industrial data

$m=2$	$PE$	$PC$	$V_{XB}$	$V_{FS}$	$V_p$
C=2	0.691	<b>0.500</b>	23.545	0.110	0.006
C=3	0.734	0.334	10.027	0.073	0.012
C=4	0.698	0.250	8.262	0.055	0.011
<b>C=5</b>	0.662	0.330	<b>0.598</b>	0.015	<b>0.340</b>
C=6	0.609	0.167	17.354	0.037	0.011
C=7	0.569	0.143	12.706	0.031	0.009
C=8	0.536	0.125	23.550	0.027	0.010
C=9	0.647	0.189	1.465	<b>0.006</b>	0.214
C=10	<b>0.476</b>	0.100	9.306	0.022	0.008

After rule-base generation and parameter learning, a 5-rule fuzzy model was obtained. The distribution of the membership functions for each input variable is represented in Fig.3. It is worth noting that the obtained fuzzy rule-based model reveals relationships between composition-process condition and tensile strength, which are consistent with metallurgical knowledge. The modeling result with RMSE (Root-Mean-Square-Error)=40.72 and 43.29 for training (1902 data) and testing (1902 data) respectively, is shown in Fig. 4. It can be seen that the final fuzzy model has a simple structure and satisfactory accuracy with good interpretability.

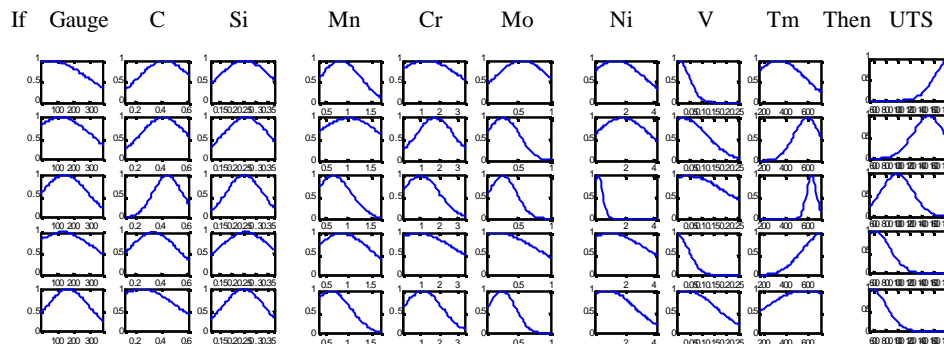


Fig. 3 Fuzzy model for Tensile Strength prediction

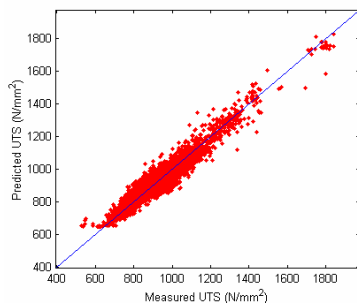


Fig.4 Model predicted TS versus measured TS

## 5. CONCLUDING REMARKS

A hierarchical fuzzy clustering approach associated with a cluster validity criterion has been proposed. This approach deals with both fuzzy clustering and cluster validity problems in a unified framework. The proposed fuzzy cluster validity index measures both compactness and separation of fuzzy  $c$ -partitions. Compared with several commonly used cluster validity indexes, the proposed validity criterion performs better in selecting the optimal number of clusters in a noisy environment. As the proposed validity measure is simply a function of the fuzzy partition matrix  $U$ , it is computationally simple but effective only in a limited range of  $m$  value ( $1.5 \leq m \leq 5$ ). Since the sub-clusters (with much smaller number than the original data) generated by the self-organizing network provide more dense and representative information for the fuzzy clustering process, the computing burden of the subsequent fuzzy clustering is reduced dramatically, especially in the situation of a large number of data. Also, the performance of fuzzy clustering with noise-contaminated data is improved due to the self-organizing network. This approach has been successfully used in alloy material property prediction. This is also a fast way to generate fuzzy models based on fuzzy clustering techniques.

## REFERENCES

- Bezdek, J. C. (1974), Cluster validity with fuzzy sets, *J. Cybernetics*, vol.3, no.3, pp.58-72.
- Bezdek, J. C. (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum.
- Dave, R. N.(1993), Robust fuzzy clustering algorithms, *Proceedings of the 2nd IEEE Int. Conf. on Fuzzy Systems*, San Francisco, pp. 1281-1286.
- Fukuyama Y. and Sugeno, M. (1989), A new method of choosing the number of clusters for the fuzzy  $c$ -means method, *Proc. 5<sup>th</sup> Fuzzy Syst. Symp.*, pp. 247-250.
- Kirshnapuram, R. and Keller, J. (1993), A possibilistic approach to clustering, *IEEE Trans. on Fuzzy Systems*, Vol.1, No.1, pp. 98-110.
- Kohonen, T. (1995), *Self-Organizing Maps*, *Springer Series in Information Science*, Vol.30, Springer-Verlag,.
- Linkenns, D. A. and Chen Minyou (1998), Hierarchical Fuzzy Clustering Based on Self-organising Networks, *Proceedings of IEEE FUZZY-98, IEEE*, Piscataway, Vol. 2, pp.1406-1410.
- Nasraoui, O. and Krishnapuram, R. (1996), An Improved *Possibilistic C-means Algorithm with finite Rejection and Robust Scale Estimation*, *Proceedings of the 5th IEEE Int. Conf. on Fuzzy Systems*, pp.395-399.
- Pedrycz, W. (1996), Conditional Fuzzy C-Means, *Pattern Recognition Letters*, Vol.17, pp. 625-631.
- Xie X. L. and Beni G. A. (1991), Validity measure for fuzzy clustering, *IEEE Trans. Pattern Anal. Mach. Intell.*, **13**(8), pp. 841-846.