On the Automatic Construction of Hierarchical Fuzzy Systems Using a Binary Interclass Separability Criterion

A. Chong[1], T.D. Gedeon[1], and L.T. Koczy[1,2]

[1]School of Information Technology
Murdoch University
South Street, Murdoch, WA, 6150
AUSTRALIA
{cchong, tgedeon}@murdoch.edu.au

[2]Department of Telecom & Telematics
Budapest University of Technology and Economics
HUNGARY
koczy@ttt.bme.hu

Abstract—This paper investigates the automatic construction of hierarchical fuzzy systems. Given a set of training data, homogeneous structures (or clusters) are searched for within the data using fuzzy c-means clustering. The clusters found are used to construct sub-rule bases in the lower hierarchical levels of the entire rule-base. For each cluster found, the proposed binary interclass-separability criterion is used to determine the suitable subspace that forms a separation surface. The meta-rules are formed using the subspaces identified. The proposed technique is validated using both artificially generated as well as real world data.

I. Introduction

THIS paper investigates the automatic construction of hierarchical fuzzy systems [1]. The hierarchical fuzzy system is based on the following idea. Often, the multi-dimensional input space $X = X_1 \times X_2 \times \ldots \times X_k$ can be decomposed into some subspaces, e.g. $Z_0 = X_1 \times X_2 \times \ldots \times X_{k0}$ $(k_0 < k)$, so that in Z_0 a partition $\Pi = \{D_1, D_2, D_3\}$ can be determined. In each D_i, a sub-rule base R_i can be constructed with local validity. The hierarchical rule base structure becomes:

R_0: if Z_0 is D_1 then use R_1
if Z_0 is D_2 then use R_2
⋮
if Z_0 is D_n then use R_n

R_1: if Z_1 is A_{11} then y is B_{11}
if Z_1 is A_{12} then y is B_{12}
⋮
if Z_1 is A_{1m1} then y is B_{1m1}

… …

R_n: if Z_n is A_{n1} then y is B_{n1}
if Z_n is A_{n2} then y is B_{n1}
⋮
if Z_n is A_{nmn} then y is B_{nmn}

The fuzzy rules in rule base R_0 are termed meta-rules since the consequences of the rules are pointers to other sub-rule bases instead of fuzzy sets. We denote Z_i as the subspace used in the rule base R_i. D_i is the i^{th} fuzzy set used in the meta-rule base R_0 and A_{ij} is the j^{th} fuzzy set used by the rule base R_i. In the sample model above, only one meta-rule base (R_0) is used. In this paper, it is sufficient to restrict our discussion to the sample model but we remark that more than one meta-rule base can be present in practice. This will be illustrated in our experiment later in section VI.

The two main advantages of the hierarchical structure are interpretability and reduced complexity. Hierarchical structured rules have fewer terms in the antecedents leading to higher interpretability compared to traditional fuzzy rules which use all input dimensions in the antecedents. The complexity of fuzzy systems can be reduced when suitable Z_0 and Π are found such that in each sub-rule base R_i the input space X_i is a subspace of $X / Z_0 = X_{k0+1} \times X_{k0+2} \times \ldots \times X_k$ [1].

The inference mechanism of the hierarchical fuzzy system has been established in [1]. Here, we consider the automatic construction of such hierarchical rule bases from a set of training data. The difficulty in construction is mainly in finding the subspace Z_0 and a suitable Π.

The main requirement of a reasonable Π is that each of its elements D_i can be modeled by a rule base with local validity. In this case, it is reasonable to expect D_i to contain homogeneous data. The problem of finding Π can thus be reduced to finding homogeneous structures within the data. This can be achieved by clustering algorithms. Section II of this paper describes how fuzzy c-means clustering can be used for this purpose.

The subspace Z_0 is used by meta-rules to select the most appropriate sub-rule base to infer the output for a given observation (i.e. system input). In section III of this paper, we formulate a criterion for determining Z_0 based on the concept of interclass separability. A preliminary hierarchical fuzzy modeling scheme is proposed in section IV. Section V describes the parameter tuning process. The experiments and conclusion are presented in sections VI and VII respectively.

II. Partition Finding With Fuzzy Clustering

One of the main tasks in hierarchical modeling is finding a partition $\Pi = \{D_1, D_2, D_3\}$ from the training data such that for each D_i, a sub-rule base R_i can be constructed with local validity. For R_i with local validity, we expect D_i to contain homogeneous data, to be found by a clustering algorithm. Among the widely used clustering algorithms, Fuzzy c-Means clustering (FCMC) [2] remains predominant in the fuzzy research literature.

Given c clusters, FCMC partitions the data $X = \{x_1, x_2, \ldots, x_n\}$ into c fuzzy classes by minimizing the within group sum of squared error objective function as follows (Eqn 1).

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (U_{ik})^m \, \| x_k - v_i \|^2, \; 1 \le m \le \infty \qquad \text{Eqn 1}$$

where $J_m(U,V)$ is the sum of squared error for the set of fuzzy clusters represented by the membership matrix U, and the associated set of cluster centers V. $\|.\|$ is some inner product-induced norm. In the formula, $\|x_k - v_i\|^2$ represents the distance between the data x_k and the cluster center v_i. The number m governs the influence of membership grades in the performance index. The necessary conditions for (Eqn 1) to reach its minimum are:

$$U_{ik} = \left(\sum_{j=1}^{c} \left(\frac{\| x_k - v_i \|}{\| x_k - v_j \|} \right)^{2/(m-1)} \right)^{-1} \forall i, \forall k, \qquad \text{Eqn 2}$$

and

$$v_i = \frac{\sum_{k=1}^{n} (U_{ik})^m x_k}{\sum_{k=1}^{n} (U_{ik})^m}, \qquad \text{Eqn 3}$$

In each iteration of the FCMC algorithm, matrix U is computed using (Eqn 2) and the associated cluster centers are computed as (Eqn 3). This is followed by computing the square error in (Eqn 1). The algorithm stops when either the error is below a certain tolerance value or its improvement over the previous iteration is below a certain threshold.

The optimal number of clusters are determined by means of a criterion, known as the cluster validity index. Fukuyama and Sugeno (FS) proposed the following cluster validitiy index [3]:

$$S(c) = \sum_{k=1}^{n} \sum_{i=1}^{c} (U_{ik})^m (\|x_k - v_i\|^2 - \|v_i - \bar{x}\|^2) \;\; 2 < c < n \qquad \text{Eqn 4}$$

where n is the number of data points to be clustered; c is the number of clusters; x_k is the k^{th} data, $\bar{x}$ is the average of data; v_i is the i^{th} cluster center; U_{ik} is the membership degree of the k^{th} data with respect to the i^{th} cluster and m is the fuzzy exponent. The number of clusters, c, is determined so that $S(c)$ reaches a local minimum as c increases. The terms $\| x_k - v_i \|$ and $\| vi- \bar{x} \|$ represent the variance in each cluster and variance between clusters respectively.

The performance of the validity index can be improved when used in conjunction with a merging index [4]. The hybrid approach proposed in [4] for finding the optimal number of clusters has two steps. In the first step, the FS index is used to find a rough estimation of the optimal number of clusters. The number is later refined in the second step by merging pairs of clusters based on the following index:

$$P(v) = \sum_{j=1}^{n} e^{-4\left\| (v - x_j) / \frac{(v_i - v_j)}{2} \right\|^2} \qquad \text{Eqn 5}$$

where x_j is the j^{th} data and v is a cluster center. For each pair of cluster centers v_i and v_j, the index (Eqn 5) is calculated for v_i, v_j, and v_m where v_m is the middle point $(v_i + v_j)/2$. If $p(v_m)$ is smaller than both $p(v_i)$ and $p(v_j)$, then the centers stay un-merged. Otherwise, they are merged.

III. Finding Subspace For Meta-rules Construction

In this section, the problem of subspace determination for meta-rules construction is discussed. In a hierarchical fuzzy system, the subspace Z_0 is used by meta-rules to select the most appropriate sub-rule base to infer the output for a given observation. Given the partition $\Pi = \{D_1, D_2, \ldots D_n\}$, a subspace Z_0 can be determined by considering its ability to separate the components D_i in the partition. The intuition is that the more separable the components, the easier the sub-rule base selection becomes. In this paper, we adapt the interclass separability index for finding the subspace Z_0. A review of the original interclass separability index often used for feature selection is presented in subsection A, and extended for hierarchical fuzzy modeling in subsection B.

A. *Interclass Separability*

The concept of interclass separability was successfully used in feature selection [5]. By feature selection, we refer to the process of identifying input variables that has significant influence to the output. Here and elsewhere in the paper, the term feature and input variable are used interchangeably. Consider a set of N input-output pairs $F = \{X; y\}$, $X = \{x_i \mid i \in I\}$ where I is the index set, x_i and y are column vectors. By deleting some features, we obtain a subspace $X' = \{x_i \mid i \subset I\}$. Suppose that the input X should be categorized into classes C_i ($i = 1, \ldots, N_c$) which possess a priori class probability p_i and the cardinality of the classes is $|C_i| = n_i$, then the criterion function for feature ranking based on the interclass separability is formulated by means of the following between-class (Eqn 6) and within-class (Eqn 8) scatter (covariance) matrices.

$$Q_b = \sum_{i=1}^{N_c} p_i \left(v_i - \bar{v}\right)\left(v_i - \bar{v}\right)^T \qquad \text{Eqn 6}$$

$$Q_i = p_i \frac{1}{n_i} \sum_{k=1}^{n_i} \left(x'_{ik} - v_i\right)\left(x'_{ik} - v_i\right)^T \qquad \text{Eqn 7}$$

$$Q_w = \sum_{i=1}^{N_c} Q_i \qquad \text{Eqn 8}$$

where

$$p_i = \frac{n_i}{N} \qquad \text{Eqn 9}$$

$$v_i = \frac{1}{n_i} \sum_{K=1}^{n_i} x'_{ik} \qquad \text{Eqn 10}$$

$$\bar{v} = \sum_{i=1}^{N_c} p_i v_i \qquad \text{Eqn 11}$$

Here, T denotes the matrix transpose. The criterion is a trade off between Q_b (Eqn 6) and Q_w (Eqn 8), often expressed as:

$$J(X') = \frac{tr(Q_b)}{tr(Q_w)} \qquad \text{Eqn 12}$$

where '*tr*' denotes the trace of a matrix. In [6] the interclass separability criterion was adapted for fuzzy modeling by generalizing Q_b and Q_i to (Eqn 13) and (Eqn 14).

$$Q_b = \sum_{i=1}^{N_c} \sum_{j=1}^{N} \mu_{ij}^m \left(v_i - \bar{v}\right)\left(v_i - \bar{v}\right)^T \qquad \text{Eqn 13}$$

$$Q_i = \frac{1}{\sum_{j=1}^{N} \mu_{ij}^m} \sum_{j=1}^{N} \mu_{ij}^m \left(x_j - v_i\right)\left(x_j - v_i\right)^T \qquad \text{Eqn 14}$$

where

$$v_i = \frac{1}{\sum_{j=1}^{N} \mu_{ij}^m} \sum_{j=1}^{N} \mu_{ij}^m x_j \qquad \text{Eqn 15}$$

$$\bar{v} = \frac{1}{N_c} \sum_{i=1}^{N_c} v_i \qquad \text{Eqn 16}$$

In [6], the set of classes C is determined by clustering the output space using fuzzy clustering algorithms such as fuzzy c-means [2]. The resulting cluster-membership degrees (μ_{ij}) are then used as weights (Eqn 13 - Eqn 15).

Sequential backward selection (SBS) is used to rank the features as follows.

1. Let $F = \{1, \ldots, N\}$ be the complete set of features.
2. For all $f \in F$
   a. Let $F = F - \{f\}$ and also update matrix X, vectors v_i and v by deleting temporarily its f^{th} row or element.
   b. Calculate matrices $Q_b(X)$, $Q_w(X)$ and determine $J(X')$.
3. Let $f' = \operatorname{argmin}_{f \in F} J(X|_f)$, that is where J attains its minimal value. Delete permanently the variable(s) f' from F the corresponding columns from X, v_i and v. Note that f' can contain more than one variable.
4. If $\{F\} > 1$ then go to step 2, otherwise stop.

The order of the deleted variables gives their rank of importance. Based on the ranking produced, the optimal number of features can be determined on a trial and error basis. Specifically, we start to build up fuzzy systems with a small number of top ranked features (one or two). The system is evaluated by a performance index. The process is repeated with a slightly increased number of features (e.g. three) until either an acceptable performance is achieved or a local optimum is reached.

With some modification, the technique can assist the determination of Z_0 in hierarchical fuzzy modeling. In the next section, the modification is presented.

B. *Binary Interclass Separability*

Consider the set of homogeneous structures (or classes) C_i ($i = 1, \ldots, N_c$) identified from a set of data for which each C_i can be modeled by a sub-rule base R_i with local validity. Given an observation, we need to select the most suitable rule base to infer the output. This job is performed by meta-rules whose output is the chosen sub-rule base (see section I).

One criterion for finding the suitable subspaces Z_i to form the meta-rules is based on the separability of the classes C_i. That is, for each C_i, the validity of the corresponding Z_i can be determined by computing its ability to separate C_i from the rest of the classes C_j $\forall j \neq i$. Let v_i be the class mean of C_i, i.e. the mean of data points in C_i, the validity criterion can be formulated by considering the separation between v_i and v_r where:

$$v_r = \frac{1}{N_c - 1} \sum_{j \neq i} v_j \qquad \text{Eqn 17}$$

This way, we deal with the separation of only two classes at a time, thus the name binary interclass separability.

Consequently, (Eqn 8) and (Eqn 13) can be simplified to (Eqn 18) and (Eqn 19) respectively.

$$Q_w = Q_i + Q_r \qquad \text{Eqn 18}$$

$$Q_b = \sum_{j=1}^{N} \mu_{ij}^m \left(v_i - \bar{v}\right)\left(v_i - \bar{v}\right)^T + \sum_{j=1}^{N} \mu_{rj}^m \left(v_r - \bar{v}\right)\left(v_r - \bar{v}\right)^T \qquad \text{Eqn 19}$$

where

$$\mu_{rj} = 1 - \mu_{ij} \qquad \text{Eqn 20}$$

$$\bar{v} = \frac{v_i + v_r}{2} \qquad \text{Eqn 21}$$

The validity of the subspace Z_i can be calculated by (Eqn 12). Using the binary interclass separability criterion, the subspace Z_i can be determined for each class C_i using an algorithm analogous to the techniques discussed in section III(A). When Z_i is found, a fuzzy set D_i can be constructed by examining the data points from the subspace Z_i.

Subsequently, a meta-rule can be formed as:

If Z_i is D_i then use R_i

IV. HIERARCHICAL FUZZY MODELING

In this section, we propose an algorithm for hierarchical fuzzy modeling. The algorithm consists of 5 steps.

1. Find homogeneous structures within the data. Perform Fuzzy c-means clustering [2] on the data. The optimal number of clusters within the set of data is determined by means of a cluster validity index (see section II).
2. A crisp partition of the data points are constructed based on the set of clusters (each with homogeneous structure), namely, for each C_i, a crisp cluster of points are determined $P_i = \{p \mid \mu_i(p) > t\}$ where t is a small threshold.
3. For each cluster C_i, rank the features according to the binary interclass separability criterion proposed in section III(B). Select n top-ranked variables to form the subspace Z_i in which the following meta-rule is constructed.

   If Z_i is D_i then use R_i

   The fuzzy set D_i can be constructed by examining the crisp partition P_i. At this stage, the number n is left as a user-input parameter. Further research is necessary to automate the selection.
4. Construct the sub-rule bases. For each crisp partition P_i, apply a feature extraction algorithm (see section III(A)) to eliminate unimportant features. The remaining features (know as true inputs) are then used by a fuzzy rule extraction algorithm to create the fuzzy rule base R_i. Here, we suggest the use of the projection-based fuzzy modeling approach [7]. We remark however, that the hierarchical fuzzy modeling scheme does not place any restriction on the technique used. If more hierarchical levels are desired, repeat steps 1 – 5 with the data points in P_i.
5. Parameter Identification. The completed hierarchical fuzzy rule-base then goes through a parameter identification process where the parameters of the membership function used in the fuzzy rules are adjusted to improve the overall performance. Section V describes the process in more detail.

V. PARAMETER IDENTIFICATION

Parameter identification is a process to tune the parameters of membership functions in the rule antecedents. The technique described in [8] is designed for trapezoidal fuzzy sets. The algorithm is as follows:

1. Set the value f for adjustment.
2. Let p^k_j be the k^{th} parameter of the j^{th} fuzzy sets.
3. Calculate $p^{k+}_j = p^k_j + f$ and $p^{k-}_j = p^k_j - f$.
   If k = 2, 3, 4, and $p^k_j + f > p^{k+1}_j$, then $p^{k+}_j = p^{k-1}$
   If k = 1, 2, 3, and $p^k_j - f < p^{k-1}_j$, then $p^{k-}_j = p^{k+1}$
4. Choose the parameter with the best performance among $\{p^{k+}_j, p^k_j, p^{k-}_j\}$ and replace p^k_j by it.
5. Go to step 2 while unadjusted parameters exist.
6. Repeat step 2 until satisfied with performance.

In [8], f = 5% of the width of the universe of discourse is used. Fig. 1 shows the parameter adjustment process.

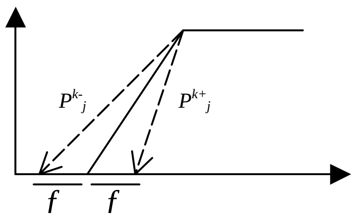

Fig. 1. Parameter Adjustment.

VI. EXPERIMENTS

In this section, we present the experiments done to validate the potential of the proposed hierarchical modeling scheme and the binary interclass separability criterion. Both artificially generated and real world data are used in the experiments. The next sub sections described the experiments.

A. Experiments with Artificial by Generated Data

A simple hybrid system is designed to generate data with a hierarchical nature and the binary interclass separability criterion is applied in conjunction with the hierarchical rule base generation scheme to the data.

The structure of the hybrid system consists of three sub-models controlled by three crisp rules illustrated as follows.

If X_1 is [0.1 0.9] then use M_1
If X_3 is [0.1 0.9] then use M_2
If X_5 is [0.1 0.9] and X_6 is [0.1 0.9] then use M_3

M_1: $y = f_1(X_2, X_4) = (X_2)^2 + (X_4)^{1.5}$
M_2: $y = f_2(X_7) = \sin(X_7)$
M_3: $y = f_3(X_8) = \cos(X_8)$

The hybrid system has 8 input variables, namely, $X_1, \ldots, X_8$. X_1, X_3, X_5 and X_6 are used by the meta-rule base to select one of the sub-models to infer the output given an observation. Different sub-models M_i use different subsets of inputs to infer the output (e.g. M_1 uses X_2 and X_4). Altogether 300 data points are input to the hybrid system to generate the corresponding output. To keep the experiment simple, the first 100 data points trigger only rule number 1, the subsequent 100 data points trigger only rule number 2 and the last 100 data points trigger only rule number 3. Random normal values are inserted in dimensions X_2, X_4, X_6, X_7.

Following the scheme in section IV, we perform fuzzy c-means clustering in conjunction with a cluster validity index on the data in the multi-dimensional input space. Three clusters were found. The cluster centers are shown in Fig. 2.

[1.52, 1.48, 1.46, 2.48, **0.50**, **0.50**, 3.49, 4.52]
[**0.52**, 1.51, 1.46, 2.51, 1.50, 1.50, 3.50, 4.51]
[1.49, 1.46, **0.52**, 2.47, 1.52, 1.52, 3.47, 4.49]

Fig. 2. Cluster Centers

The ranking from section III(B) is shown in Table 1. The dominating variables in the table are based on our prior knowledge of the original model. It can be observed from the results that the criterion produces correct rankings in the sense that the relevant features in the crisp rules are ranked highest in all three clusters.

Table 1

EXPERIMENTAL RESULTS

Cluster	Dominating Variable	Feature Ranking
C_1	X_5, X_6	**5 6** 1 3 2 4 7 8
C_2	X_1	**1** 5 6 3 2 4 7 8
C_3	X_3	**3** 5 6 1 2 4 7 8

In the following, the experiment is repeated to simulate a less than ideal situation where the number of clusters found does not match the desired number. Experiment II and III were performed by fuzzy clustering the data into 5 and 12 clusters respectively. The clusters' dominating variables as well as the ranking are shown in Table 2 and Table 3. The criterion correctly gives top ranking to the dominating variables.

Table 2

EXPERIMENTAL RESULTS II

Cluster	Dominating Variables	Feature Ranking
C_1	X_5, X_6	**5 6** 1 3 8 4 2 7
C_2	X_3	**3** 5 6 1 2 8 4 7
C_3	X_5, X_6	**5 6** 1 3 8 4 2 7
C_4	X_1	**1** 5 6 3 2 4 7 8
C_5	X_1	**1** 5 6 3 2 4 7 8

Table 3

Experimental Results III

Cluster	Dominating Variables	Feature Ranking
C_1	X_1	**1** 3 5 6 2 4 7 8
C_2	X_3	**3** 1 5 6 2 4 8 7
C_3	X_1	**1** 3 5 6 2 4 7 8
C_4	X_3	**3** 1 5 6 2 4 8 7
C_5	X_5, X_6	**5 6** 1 3 8 4 2 7
C_6	X_5, X_6	**5 6** 1 3 8 4 2 7
C_7	X_3	**3** 1 5 6 2 4 8 7
C_8	X_1	**1** 3 5 6 2 4 7 8
C_9	X_1	**1** 3 5 6 2 4 7 8
C_{10}	X_1	**1** 3 5 6 2 4 7 8
C_{11}	X_5, X_6	**5 6** 1 3 2 4 7 8
C_{12}	X_3	**3** 1 5 6 2 4 7 8

Next, we apply the hierarchical modeling scheme to the data.

The meta-rules generated for the three clusters in Fig. 2 are shown as follows where the vector denotes the four characteristic points of a trapezoid.

1. If X_5 is [0.14 0.29 0.75 0.84] and
   X_6 is [0.14 0.29 0.75 0.84] then *use* R_1
2. If X_3 is [0.10 0.38 0.60 0.90] then *use* R_2
3. If X_1 is [0.10 0.34 0.65 0.90] then *use* R_3

It can be seen from the results that the meta-rule base generated is substantially similar to the original meta-rules used in the artificially generated hybrid model.

Next, the data points in each crisp cluster P_i created at step 2 are used to create a sub-rule base R_i. In step 4, it is suggested that a feature selection algorithm is applied on the data before the rule extraction process. In this experiment, we adopt the feature selection technique described in section III(A). Table 4 shows the feature ranking results for the data points in each crisp cluster. The results matches our expectation based on the hybrid model. Subsequently, a sub-rule base can be constructed using the true inputs identified to model each of the sub-models designed in the experiment. The construction of the sub-rule base can be performed by using any fuzzy rule-extraction algorithm in the literature [9].

Table 4

Feature Ranking

Cluster	Dominating Variables	Feature Ranking
P_1	X_8	**8** 7 3 5 6 2 1 4
P_2	X_7	**7** 4 5 6 2 1 3 8
P_3	X_2, X_4	**2 4** 1 5 6 7 8 3

B. Experiment With Real-World Data

The data used is a set of benchmark data in reservoir characterization, obtained from a real reservoir. The objective is to develop an estimator to predict porosity (PHI) from well logs.

The well logs available are: Depth, GR (Gamma Ray), RDEV (Deep Resistivity), RMEV (Shallow Resistivity), RXO (Flushed Zone Resistivity), RHOB (Bulk Density), NPHI (Neutron Porosity), PEF (Photoelectric Factor) and DT (Sonic Travel Time). Normalised data [0, 1] is used.

There are altogether 633 rows of data. Since accuracy and the generalization ability are not the main concern of this study, the same set of data is used for training as well as testing. The goal of the experiment is to verify the practicality of the hierarchical model by constructing a hierarchical fuzzy system with reasonable accuracy and good interpretability out of real world data. To ensure that the model has reasonable accuracy, the mean square error has been used as a performance index:

$$PI = \sum_{i=1}^{m} (y^i - \hat{y}^i)^2 / m \qquad \text{Eqn 22}$$

Where m is the number of data, y^i is the i^{th} actual output and $\hat{y}^i$ is the i^{th} model output.

In step one, 2 clusters were found within the data using FCMC in conjunction with the validity index. After the creation of a crisp partition in step two, the binary interclass separability index is applied to the data. The ranking shows that the input variable 'Depth' is most able to separate the two clusters. This coincides with our understanding of the petroleum data. In general, the rocks at different depths exhibit different behavior. For example, porous rocks can collect hydrocarbon but must be capped by non-porous nodes to retain them through time.

Using the variable 'Depth', the fuzzy meta-rules in Fig. 3 are produced. For each crisp class of points, unimportant features are eliminated through the use of the interclass separability index (section III(A)) and a sub-rule base is constructed with the remaining features. The two sub-rule bases generated are shown in Fig. 4 and Fig. 5. After 4 iterations of parameter tuning, the performance index is 0.023, which is satisfactory.

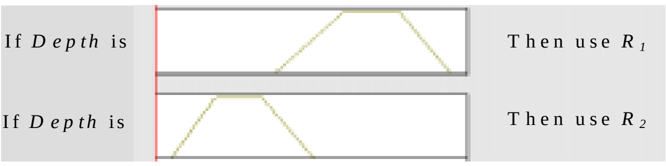

Fig. 3. Fuzzy Meta-rule Base for Petroleum Data.

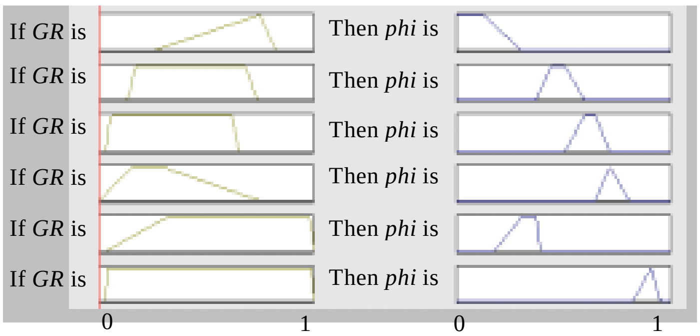

Fig. 4. Sub-rule Base R_1

VII. Conclusion

The automated construction of hierarchical fuzzy systems has been investigated. An algorithm for hierarchical fuzzy rule extraction has been proposed. Firstly, homogeneous structures (or clusters) are searched for within the training data using fuzzy clustering algorithms. The clusters found are used to construct sub-rule bases in the lower hierarchical level of the entire rule-base. For each cluster found, the proposed binary interclass-separability criterion is used to determine the suitable subspace that forms a separation surface. The subspaces found are then used to formulate meta-rules.

The proposed algorithm is validated using both artificially generated and real world data. In subsequent work, some implementation improvements will be investigated.

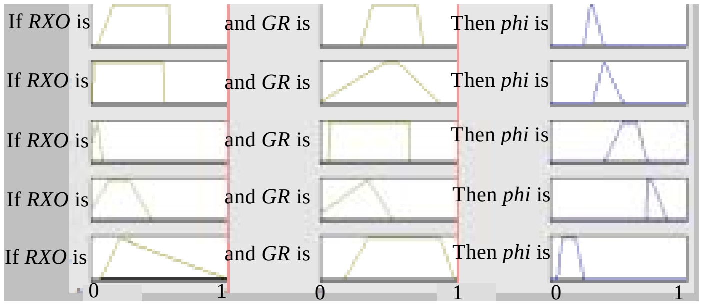

Fig. 5. Sub-rule Base R_2.

References

[1] Koczy, L.T. and K. Hirota. Approximative inference in hierarchical structured rule bases. in *Fift IFSA World Congress*. 1993. Seoul: International Fuzzy Systems Association: p. 1262-1265.

[2] Bezdek, J.C., *Pattern Reconition with Fuzzy Objective Function Algorithms*. 1981, New York: Plenum Press.

[3] Fukuyama, Y. and M. Sugeno. A new method of choosing the number of clusters for fuzzy c-means method. in *Proceedings of the 5th Fuzzy Systems Symposium*. 1989: p. 247-250.

[4] Chong, A., T.D. Gedeon, and L.T. Koczy. A Hybrid Approach for Solving the Cluster Validity Problem. in *submitted to 14th International Conference on Digital Signal Processing*. 2002. Greece: p. 1207-1210.

[5] Devijver, P.A. and J. Kittler, *Pattern Recognition: A Statistical Approach*. 1982, London: Prentice Hall.

[6] Tikk, D. and T.D. Gedeon. Feature ranking based on interclass separability for fuzzy control aplication. in *Proceedings of the International Conference on Artificial Intelligence in Science and Technology (AISAT'2000)*. 2000. Horbat: p. 29-32.

[7] Chong, A., T.D. Gedeon, and L.T. Koczy. Projection Based Method for Sparse Fuzzy System Generation. in *2nd WSEAS Int. Conf. on Scientific Computation and Soft Computing*. 2002. Crete: p. 321-325.

[8] Sugeno, M. and T. Yasukawa, A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1993. **1**(1): p. 7-31.

[9] Guillaume, S., Designing Fuzzy Inference Systems from Data: An Interpretability-Oriented Review. *IEEE Transactions on Fuzzy Systems*, 2001. **9**(3): p. 426-443.