Comparison of Different Strategies of Utilizing Fuzzy Clustering in Structure Identification

Kemal Kiliç(*) [a]

[a]*FENS, Sabanci University, Istanbul, Turkey, kkilic@sabanciuniv.edu*

Özge Uncu [b]

[b]*Simon Fraser University, Burnaby, BC, Canada*

I. Burhan Türksen [c]

[c]*TOBB University of Economics and Technology, Ankara, Turkey*

Abstract

Fuzzy systems approximate highly nonlinear systems by means of fuzzy "if-then" rules. In the literature, various algorithms are proposed for mining. These algorithms commonly utilize fuzzy clustering in structure identification. Basically, there are three different approaches in which one can utilize fuzzy clustering; the first one is based on input space clustering, the second one considers clustering realized in the output space, while the third one is concerned with clustering realized in the combined input-output space. In this study, we analyze these three approaches. We discuss each of the algorithms in great detail and offer a thorough comparative analysis. Finally, we compare the performances of these algorithms in a medical diagnosis classification problem, namely Aachen Aphasia Test. The experiment and the results provide a valuable insight about the merits and the shortcomings of these three clustering approaches.

Key words:
Fuzzy System Modelling, Medicine, Knowledge Acquisition, Data Mining, Structure Identification.

1 Introduction

A system may be considered as a set of interrelated elements structured in such a way to accomplish a common goal. A model is the model builder's description of the system in order to analyze its behavior. Therefore, "system modeling"

is an essential step in decision-making processes in order to explain, predict and control a system. Crucial information, necessary for building a realistic model, is usually hidden in the historical data. Better tools for analyzing the data lead to better modeling of the system, hence, better solutions for the decision problems. Quantitative methods, i.e., statistics, optimization and simulation, as well as soft computing techniques are commonly used for this purpose. In the decision making process, one often needs to introduce soft computing techniques in order to understand the structure and the behavior of a system that is highly nonlinear and highly uncertain. Amongst the soft computing techniques, fuzzy system modeling provides valuable knowledge to the decision maker in terms of linguistic (therefore easily comprehensible) fuzzy if-then rules that associate the inputs to the outputs. In particular, fuzzy set theory provides the necessary framework to handle further uncertainty such as the imprecision associated with the data in the modeling exercises.

In fuzzy system modeling, the nonlinear relations in the data are approximated by means of fuzzy if-then rules. In earlier approaches, the fuzzy if-then rules were determined a priori from other sources such as experts' knowledge. However this methodology is highly subjective, i.e., the fuzzy if-then rules usually change from expert to expert, even the same expert may suggest different rules at different times. Therefore, there is a growing research domain on modeling approaches for objective identification of the structure in the data in terms of fuzzy if-then rules (3)(4)(5)(6)(8)(9)(10)(12)(14)(13)

Hence data analysis (and/or data mining) became one of the basic steps of fuzzy system modeling. Data consist of objects that are defined in terms of some attributes. The overall goal of data mining is to find the structure of the data in terms of the relationships identified in a rule structure. The data can be viewed as a collection of ND (number of data) objects, where each object is represented by means of number of variables (NV) attributes. However, unless the structure that is hidden in the system is identified, the data provides very little information. Hence, the objective of the data analysis is to bring the hidden structure to the surface.

One of the advantages of fuzzy system modeling is the fact that it reduces the complexity of the data by using information granules and presents the data to the user in the form of perceivable fuzzy rules. Therefore, it reduces the complexity based on abundance of information. Furthermore, fuzzy system modeling allows the formation of fuzzy granules that handle vagueness of the concepts. This is a more realistic approach in a world where there are all shades of gray between black and white.

Fuzzy system modeling consists of two stages, namely the "*system identification*" stage and the "*fuzzy reasoning*" stage. The *system identification* can be informally described as the identification of the hidden rules and the relations

by incorporating the fuzzy set theory and the notions of membership gradation, approximation and similarity. In this stage, the significant input variables are determined, the fuzzy if-then rules are generated and the parameters of the model, such as the number of clusters, the level of fuzziness, the operators to be used in the reasoning, etc., are selected. The second, *fuzzy reasoning,* is the methodology to be used to infer new knowledge from the identified rule base in case of partial agreement is encountered.

Various approaches have been developed to date for these two stages of fuzzy system modelling. Generally speaking, these algorithms can be classified into three broad approaches in terms of the structure of the consequents in the fuzzy if-then rules that they generate. These are namely, the Takagi-Sugeno-Kang (11) type rule structure where the consequents are expressed as a linear combination of weighted input variables; Mamdani type fuzzy rules as later yield to famous Sugeno-Yasukawa modelling (10) in which the consequents are fuzzy sets; and the simplified fuzzy model, i.e. Mizumoto type rule structure, in which the consequents are constants(7). The focus of this research is limited with the Mizumoto type fuzzy rule structures which is actually a special case of the other two particularly for the problems where the consequents are classes (categories).

In the fuzzy data mining literature, more emphasize is given to develop algorithms that yields Mamdani type fuzzy rules. The main reason for this trend is the fact that these rule types are more descriptive then the TSK rules, since the rules themselves are based on natural language rather than mathematical functions. Generally speaking, the system identification stage of these algorithms consist of three phases; first phase is the *structure identification phase,* in which the significant inputs are determined, fuzzy membership functions are obtained and number of rules are determined; second phase is the *parameter tuning phase,* in which the parameters that will be utilized during the inference stage is determined; and the last phase is the *model validation phase,* which tests the model based on its accuracy (15).

During the *structure identification phase* the fuzzy membership values can be identified based on three different strategies (or approaches) with respect to how fuzzy clustering is utilized. Firstly, we can cluster the output space and obtain the fuzzy membership functions based on the projections of the output clusters onto the input space. Secondly we can go the other way around, that is to say first cluster the input space and project the input clusters to the output space. Or finally we can cluster the input and output space altogether and then project the multi dimensional clusters to each one of the two spaces.

Each strategy has advantages and disadvantages. The objective of this paper is to discuss the merits and shortcomings of these three approaches and compare their performances. At the moment, a comparison is not available in the

literature. However, the researchers need to be aware of the advantages and disadvantages of these approaches and make their decisions accordingly. Furthermore, we will discuss and provide some specific insight requirements for each strategy in order to handle the implementation problems. For this purpose, we will review three algorithms from the literature. Each of these three algorithms utilizes different clustering strategies that is mentioned above. We will also provide a comparison of these algorithms in terms of their predictive performances in a classification problem. Note that each strategy utilized in the earlier steps of *structure identification* leads to the necessity of developing different steps for the remaining of the algorithm. Therefore, the performance calculations need not be conclusive since *everything else* doesn't stay the same. However, such a comparison might provide some insights about the predictive performance of these strategies and their applicability to the classification problems.

In the following section, we will first introduce the notation that will be used in this paper and later provide more details on different fuzzy if-then rule structures. In the third section we will summarize the fuzzy system modelling algorithms that will be used in the analysis. Section 4 will be the part where we conduct an experimental analysis based on a medical classification problem. In this section, we will analyze the results and discuss the advantages and the disadvantages of the algorithms. We will conclude the paper with our final remarks.

2 Fuzzy If-Then Rules Structures

The following mathematical notation is used in the paper.

Let $X_1, X_2, ..., X_{NV}$ fuzzy linguistic variables in the universe of $U_1, U_2, ..., U_{NV}$, respectively, and Y be a fuzzy variable in the universe of discourse V. We will use j as the index for input variables, i.e., $j = 1, .., NV$.

Let R_i be a fuzzy relation (i.e., fuzzy rules) in $U_1 \times U_2 \times ... \times U_{NV}$ .We will denote the number of rules with c. We will use i as the index for the rules, i.e., $i = 1, .., .c$.

Each fuzzy linguistic variable can be partitioned into fuzzy sets called fuzzy linguistic labels. We will denote these fuzzy sets with A_{ij}, the fuzzy linguistic label of the jth fuzzy input variable associated with i^{th} fuzzy rule.

Let $x_k = [x_{k,1}, ..., x_{k,NV}]$ denote the input vector of the kth data , where $k = 1, ..., ND$ and y_k is the output of the kth data.

In general, a fuzzy if-then rule bases has the following structure:

$$R := ALSO_{i=1}^{c} IF\ antecedent_i\ THEN\ consequent_i \tag{1}$$

In the fuzzy system modelling (FSM) method proposed by Takagi-Sugeno-Kang (TSK) (11), the consequent part of fuzzy rules are represented by using a linear function of input variables. Thus, the rule base in TSK method can be represented as follows:

$$R := ALSO_{i=1}^{c} IF\ antecedent_i\ THEN\ y_i = a_i x^T + b_i \tag{2}$$

where $x = [x_{k,1}, ..., x_{k,NV}]$ is the input data vector, $a_i = [a_{i,1}, ..., a_{i,NV}]$ is the regression line coefficient vector associated with the ith fuzzy rule, a_{ij} is the regression line coefficient in i^{th} fuzzy rule associated with j^{th} input variable and b_i is the scalar offset of regression line in i^{th} fuzzy rule. The antecedent part of the rule is as follows;

$$antecedent_i = AND_{j=1}^{NV} x_j \in X_j\ isr\ (is\ related\ to)\ A_{ij} \tag{3}$$

A major problem with the TSK structured fuzzy rule bases is the fact that the determination of the parameters associated with this model is computationally costly and the obtained rules are hard to interpret.

On the other hand, in Mamdani type approaches the consequents are fuzzy sets (10). This approach has the advantage of being more descriptive and easier to implement. However, in general, this approach suffers in terms of the predictive performance because of some structural misrepresentations with some of the existing algorithms.

A typical fuzzy rule base in Sugeno-Yasukawa like algorithms is as follows;

$$ALSO_{i=1}^{c} IF\ Antecedent_i\ THEN\ y_i = B_i \tag{4}$$

where $antecedent_i$, has the same structure as provided in Equation 3. Note that $A_{i,j}$ and B_i 's are fuzzy sets and c denotes the number of rules.

The third type of fuzzy rule structure is known as the simplified fuzzy rule, or Mizumoto type rules (7). In this rule base structure, the consequent is a scalar. Thus, the rule base proposed by Mizumoto can be formulated as follows:

$$ALSO_{i=1}^{c} IF\ antecedent_i\ THEN\ y_i = b_i \tag{5}$$

again $antecedent_i$, has the same structure as provided in Equation 3 and $A_{i,j}$ are fuzzy sets, and b_i is a scalar and c denotes the number of rules.

Note that Mizumoto type rules are actually a special version of both Mamdani type rules and TSK type rules. For example, in a classification problem in which the consequent fuzzy sets are actually scalar Mizumoto and Mamdani rules would have the same structure. For a TSK consequent where the regression line coefficient vector is a null vector, TSK and Mizumoto rules would be equivalent.

In fuzzy control literature more interest is given to TSK type rules. However as we have stated above, these rules are not descriptive and they are harder to obtain. The determination of the optimal regression line coefficients, i.e. parameter tuning, is costly in terms of computational complexity. Therefore, in data mining applications Mamdani type fuzzy if-then rules are preferred more, since they provide descriptive information which is invaluable to the users in many applications. Even though users are usually experts of the field, Mamdani type fuzzy if-then rules might provide new relations, which they couldn't notice earlier. For classifications problems the outcome is naturally a scalar. Hence, the simplified rule structure, being a special version of Mamdani type fuzzy if-then rules where the fuzzy consequents are approximated with a scalar, is more applicable.

3   Algorithms of Fuzzy Modelling

In the literature, fuzzy clustering is extensively utilized at the *system identification*, particularly in the *structure identification* phase. There are three different alternative strategies for incorporating the fuzzy clustering at this phase. First way of incorporating fuzzy clustering, as proposed originally by Sugeno-Yasukawa (10), is based on clustering first the output space. The relation of the input variables with the output is obtained after the projection of the output clusters onto input space. A second approach is clustering the *NV* dimensional input space, projecting them onto each input variable and relating the output variables to each input clusters based on the degree of possibility (3). A third possible approach is clustering the *NV+1* dimensional space, i.e. input and output space together, and projecting the obtained clusters onto each variable in order to obtain the fuzzy if-then rules(14) .

Each strategy is based on some sort of common sense. In the first approach the idea is grouping the input space data vectors that yield to similar outcomes. The association rules are obtained between these groups and the output clusters. In the second approach, the idea is grouping the input space data vectors that are similar to each other, and later trying to associate different outcomes

6

to these groups. The third approach assumes that the distinction between the input space and output space features are somewhat artificial. Therefore, it clusters all of the data vectors based on both their input space attributes and the output space attribute. Later, it aims to develop the association rules among these two spaces.

In this paper we discuss the implementation problems and their solutions proposed in the literature and provide a comparison of their predictive performances for these three strategies. For this purpose we selected three different algorithms proposed in the literature. Each one of them utilizes a different strategy of incorporating the fuzzy clustering as discussed above. First algorithm is based on the first approach, i.e., the output clustering. Sugeno and Yasukawa are the first researchers that proposed a fuzzy system modeling algorithm that is based on output clustering. We will present briefly, the original Sugeno-Yasukawa (S-Y) approach and discuss some problems associated with it. A modified algorithm (M-A) (6) that addresses some of these problems will be provided and will be used in the analysis. The second algorithm is based on input data clustering (3). Finally we will discuss the major points of the third algorithm, which is based on $NV + 1$ dimensional clustering (14) and generates Mizumoto type rules.

Unfortunately we will only provide the major points of these algorithms because of the limited space, and would kindly requests the readers to refer to the original papers for further details. However we will provide the necessary details in other steps (other than the structure identification phase) of the *system identification* and *fuzzy reasoning* stages so that the reader might comprehend the differences of the algorithms fully.

3.1   Sugeno-Yasukawa Approach and the Modified Algorithm

Sugeno and Yasukawa (10) proposed an algorithm in order to determine the fuzzy if-then rules from the historical data. There are four main steps of the algorithm. First three steps are part of the *structure identification* stage and the final step is the *fuzzy reasoning* stage. First step is clustering the output variable. This is achieved by the well-known Fuzzy C-Means (FCM) algorithm proposed by Bezdek (2). Next step is to determine the significant input variables with a myopic neighborhood search algorithm. Third step is input membership assignment, i.e., constructing the antecedent part of the fuzzy rules. This is achieved by projecting the output membership degrees onto significant input variables. The fourth step is the fuzzy inference as stated earlier.

One of the major sources of the problems associated with the S-Y algorithm is the way it handles the input membership assignment. The problem of Sugeno-

Yasukawa algorithm is one of the common problems that should be addressed by any output clustering approach based fuzzy system modeling algorithm. Therefore, first we will discuss this step in more detail.

After the output space is clustered, i.e. the membership degrees $(\mu_{B_i}(y))$ are determined, these clusters are projected onto each input variable (in order to form the $A_{i,j}$'s), one by one. That is to say, the membership degree of the value of the kth data vector's j^{th} input variable, i.e., $\mu_{A_{ij}}(x_{kj})$ is set to be equal to $\mu_{B_i}(y_k)$ for the ith fuzzy rule. Next a trapezoidal fuzzy set is fitted to these membership degrees for each input variable in order to construct the $A_{i,j}$'s. Note that in this approach, each output cluster corresponds to a single fuzzy if-then rule. Therefore, if there are c output clusters then there are also c fuzzy rules. The major drawback of this approach is the fact that while projecting the output fuzzy clusters onto input space, the natural ties among the input variables are broken and each input variable is partitioned separately. This approach neglects the possible correlations among the input variables. The modified algorithm (M-A) (6) addresses this problem and solves it by partitioning the input space into NV-dimensional clusters.

In the modified approach, after the output space is clustered, the output clusters are projected onto NV-dimensional input space. Therefore, in the M-A, a rule structure, which has NV-dimensional single antecedent, is proposed unlike the original S-Y algorithm. The fuzzy if-then rule has the same structure presented in Equation 4 However the $antecedent_i$ does not any more have the structure provided in Equation 3, but is as follows:

$$Antecedent_i = x_j \in X_j \ isr \ (is \ related \ to) \ A_{ij} \qquad (6)$$

where A_i is an NV-dimensional fuzzy set. Such a rule structure keeps the natural ties among the input variables. Another advantage of the M-A is the fact that it does not assume any pre-specified shape of membership functions such as triangular, trapezoidal, ...etc. This is important because in the Sugeno-Yasukawa algorithm, fitting a pre-specified curve or a line to the projected data points is usually a source of misrepresentation. Furthermore, in many real life cases unimodal and convex fuzzy set assumption of Sugeno-Yasukawa does not hold.

Recall that, in S-Y algorithm, output clustering is realized by Bezdek's (2) Fuzzy C-Means (FCM) algorithm. However there are various problems that are associated with the application of this algorithm to fuzzy *structure identification* phase. These problems are namely, the problems of harmonics, the problems with the boundary fuzzy sets, the problems associated with the classification and the more general problem of cluster validity, i.e. determination of number of clusters (c) and level of fuzziness (m) (6). The main reason for

the first three problems is the fact that, whenever we can order the labels of the fuzzy sets, which is usually the case in single dimensional clustering (such as small, medium, etc.), we shouldn't allow a data point to be member of more than two consecutive clusters. Otherwise, the classification of the intermediate values results in logically incorrect cases where a larger data point has a lower membership degree to a fuzzy set representing a smaller cluster, than a smaller data point. A solution to the problem may be achieved with a perspective that limits the assignment of each point to only two consecutive fuzzy sets. This may be achieved by assigning linear membership functions to each output cluster (e.g. grid based clustering with triangular fuzzy sets). In terms of the more general cluster validity problem, numerous cluster validity indices are proposed in the literature, which often optimize pre-specified functions. However, the value of c, that optimizes these functions is not necessarily the one, which optimizes the predictive performance. Therefore a supervised approach, based on modeling error minimization is offered as an alternative in the modified algorithm.

In data analysis one of the most important steps is determination of the significant input variables. S-Y proposes a neighborhood search algorithm; choose an input variable one at a time that yields the best training error. Hence, in S-Y algorithm, an input variable is either significant or not. However this is a non-fuzzy way of thinking. Some input variables may be more significant than others. The modified algorithm introduces a fuzzy learning based algorithm that determines the significance degrees of input variables. That is to say the proposed algorithm fuzzifies the concept of "significant inputs" to "significance degrees of the inputs".

For the fuzzy reasoning phase, there is a problem with the M-A rule structure, particularly in the determination of the degree of firing step during the inference. The original S-Y algorithm determines the degree of firing by separately determining the membership degree of the data to each input variable and by conjunction of these membership degrees. However, the NV-dimensional input rule structure does not fit to this frame. Therefore, a new algorithm is developed in order to determine the degree of firing which is based on a $k - NN$ algorithm.

Readers may find more details of the Sugeno-Yasukawa algorithm in (10)and modified algorithm in (6).

3.2   Castellano et al. Approach

Castellano et al. (3) suggests an alternative methodology that is defined by three major phases. The first phase is clustering the NV-dimensional input

variables space. By this way they also aim to capture the multidimensional relationships among available data. Number of multidimensional clusters is referred to as number of prototypes (NP). Later the multidimensional *prototypes* are projected on each dimension, where they are clustered into a number of one-dimensional clusters per variable. Number of fuzzy sets (NS) per dimension is a chosen a value which might be different (i.e., "less") than NP. Hence at the end of the first phase, there is NP multidimensional clusters, and NS single dimensional clusters per input variable.

The second phase utilizes the information provided by the first phase and constructs the antecedents of the fuzzy if-then rules. The fuzzy relations are formed as a Cartesian product of one-dimensional fuzzy sets and expressed as conjunction of linguistic labels. In order to avoid the combinatorial explosion, only those relations that represent the multidimensional prototypes are retained, while all others are discarded. At the end of the second phase, we obtain the antecedents of the at most NP fuzzy rules, i.e., the number of clusters (NP) obtained in the first step bounds the number of rules.

The third and final phase of the algorithm is obtaining the consequent part of the fuzzy rules. A fuzzy relation is obtained by assigning a possibility measure based on weighted occurrences of each output class. That is to say, among the set of data vectors that satisfies the antecedent parts of each fuzzy relation that is obtained after the second stage, the weighted occurrences of each class is obtained, and these occurrences becomes the consequent part of the fuzzy rules. Hence the rule structure of this approach is slightly different than the Mizumoto's rule structure provided in Equation 5. Suppose there are M different output classes, in this case the obtained fuzzy relations would have the following structure;

$$ALSO_{i=1}^{c} IF\ Antecedent_i\ THEN\ OR_{i=1}^{M}\ y_i = b_m\ with\ v_{i,m} \qquad (7)$$

where b_m's are possible output classes, and $v_{i,m}$'s are the possibility measures representing the weighted occurrences of the mth output class associated with the ith fuzzy rule.

The inference is achieved similar to the methodology described above. First a degree of similarity between the data vector and the antecedents are obtained. This similarity is used as the degree of firing of each rule. The overall outcome is obtained from the weighted (where the weights are the degree of firing) summation of the consequents.

Note that this algorithm suffers similar problems with the original Sugeno-Yasukawa approach. First of all even though the original fuzzy clustering is over the NV-dimensional input space, after the projection onto each input

variable separately, the natural ties in a data vector are lost. Furthermore, this algorithm also assumes convex and unimodal fuzzy sets, which may not be a valid assumption in many circumstances as stated earlier.

Further details of the algorithm are provided in (3).

3.3 Uncu and Turksen Approach

Uncu and Turksen (14) proposed to cluster the NV+1 dimensional data, i.e., augmented input variables and the output variable, by executing the Fuzzy C-Means algorithm with a range of c (number of clusters) and a range of m (level of fuzziness) among a candidate set. Later the obtained clusters centers are projected onto NV-dimensional input space and the corresponding memberships are projected onto output space in order to be able to calculate the center of gravity of the induced output fuzzy sets. Note that, the number of fuzzy if-then rules is also equal to the number of clusters as it was the case in the Sugeno-Yasukawa (or the modified version of it). Furthermore, the possible relationships between input variables are not broken by keeping the antecedent side of the rules in multidimensional space rather than by projecting the cluster centers onto each input variable axis separately.

The major distinctive feature of Uncu and Turksen approach is that it builds the system model by using a range of level of fuzziness values of rather than by using a single level of fuzziness value in the fuzzy clustering step. The authors considered the uncertainty in selecting the learning parameters as another source of uncertainty in model structure identification and build discrete type 2 fuzzy system models based on that idea. This discrete type 2 fuzzy system model structures can be considered as a collection of type 1 fuzzy system models identified for different level of fuzziness value. Thus, this method does not require the user to select a singleton level of fuzziness value. The best number of clusters, i.e. c, is selected based on the training error as in (6).

In the *structure identification* phase, the NV+1 dimensional clusters centers that are obtained by FCM algorithm are independently projected onto NV-dimensional input space. Thus, if we assume $v_i = (v_{i,1}, v_{i,2}, ..., v_{i,NV}, v_{i,NV+1})$ as the ith cluster center identified by FCM algorithm, the cluster center of the antecedent in i^{th} fuzzy rule can be written as $v_i^{inp} = (v_{i,1}, v_{i,2}, ..., v_{i,NV})$ where v_i^{inp} is the ith input cluster (i.e., v_i^{inp} is the cluster center of the antecedent fuzzy set associated with i^{th} fuzzy rule). The cluster center of the consequent of the ith fuzzy rule is calculated by projecting the membership values of ith $NV+1$ dimensional cluster on the output space and by taking the center of gravity of the induced output fuzzy set. Hence the obtained rules are Mizumoto type rules with scalar outputs. In addition to antecedent cluster center and output

center of gravity parameters, the approach proposed in (14) also stores an m-lookup table in which each training input data vector is associated with an m value that gives the least error for the corresponding input data vector.

Uncu and Turksen (14). also suggest a distance based similarity measure in order to determine the degree of firing for each rule. Since, the local optimal membership values are obtained by using FCM algorithm, the authors proposed to use the membership function formulation of the FCM algorithm instead of using another method of calculating degree of fire. When a new test input data vector is presented to the model, first the closest training input data vector is chosen with respect to a distance measure. Then, the type 1 fuzzy system model corresponding to the *m value* associated with the selected training input data vector is selected. After this type reduction step, the inference method proposed by Delgado et al. (5) is employed in order to calculate the model output.

Further details of the algorithm are available in (14). Note that Uncu and Turksen used a different approach in (13) to select the significant input variables.

4   Experimental Analysis and Discussion

The performances of the three algorithms are tested with the publicly available Aachen Aphasia Test (AAT) (http://fuzzy.iau.dtu.dk/aphasia.nsf/PatLight). Aphasia is the loss or impairment of the ability to use or comprehend words – often a result of stroke or head injury. Data of 256 aphasic patients, treated in the Department of Neurology at the RWTH Aachen, were collected in a database since 1986. The database consists of the clinical diagnosis as well as the diagnosis of the aphasia type and AAT profiles. Additionally 146 patients were analyzed regarding their anatomical lesion profiles with computed tomography. The original AAT has 30 attributes, including AAT scores, nominal values and images of the lesion profiles. The full detail of the data set can be found in (1).

We tried to analyze the performances of the three fuzzy system modelling algorithms discussed above in terms of the classification accuracy. Castellano et al. conducted some analysis in (3) for the AAT data set. Therefore, we decided to implement the same experimental design suggested by Castellano et al. in order to make a better comparison of the three algorithms.

The data is preprocessed and only 146 cases corresponding to the four most common aphasia diagnoses were selected in (3). These diagnoses are: Broca (motor or expressive aphasia), Wernicke (sensory or receptive aphasia), Anomic

12

AAT Scores	Description
P1	Articulation and prosody (melody of speech)
P5	Syntactic structure (structure of sentences)
N0	Repetition
C1	Written Language (reading loud)

Table 1

The AAT scores used in the analysis

$M - A$	Castellano et al.	U.T
%89.8	%78.4	%83.6

Table 2

The classification performances of the three algorithms in terms of percentage of successful classifications

(difficulties in retrieval of words) and Global (total aphasia). The authors selected AAT scores suggested in (3) for the analysis; hence, we also used the same attributes in the analysis. The selected AAT scores are tabulated in Table 1. To sum up, final database consisted of 146 cases, 4 attributes and the diagnoses. 20-fold stratified cross validation strategy is used in the experiments as suggested by the authors.

Castellano et al. conducts the experiments for different number of NP and NV values. The results obtained by Castellano et al. are tabulated in table 2 of (3). The classification error varies between 52.5% and 12%. That is to say the most successful results is 88%, where as the worst is 47.5%. The average percentage of success is 78.4 % for Castellano et al.

M-A classified correctly 131 cases (misclassified only 15 cases) in the database of 146 data vectors yielding approximately 89.8% of success rate (or 10.2 % of classification error). Note that this result is better than even the best value obtained from the Castellano et al. approach.

The Uncu and Turksen algorithm that is based on NV+1 dimensional clustering misclassified 24 cases out of 146 data vectors. This result corresponds to 83.6 % success rate (or 16.4 % classification error).

The classification performances of the three algorithms are tabulated in Table 2.

As we have mentioned earlier the results are by no means conclusive leads to a final remark about the predictive performance of the different strategies of incorporating the fuzzy clustering. First of all a single data set is not enough for such a comparison. Furthermore the algorithms that are used also differs in many aspects with each other. Even though some of these aspects are nec-

essary steps implied because of the selected strategy, there is still considerable difference that are based on the different preferences (e.g. the determination of significant inputs differs among the algorithms, etc.) However one can at least conclude that the strategies are all applicable with acceptable degree of success to classification problems.

In the rest of this section we would like to discuss our observations about these fuzzy system modelling algorithms and raise some issues that should be kept in mind while deciding, which fuzzy clustering strategy should be incorporated.

One of the advantages of the output clustering approach compared to the other two approaches is that the fuzzy rules obtained are *homogeneous* in terms of the outputs. Therefore it provides an answer to the question of "What type of inputs are associated with such an outcome?" which might be helpful to the users. On the other hand, the other two strategies have a danger of resulting fuzzy if-then rules with coinciding consequences, hence can't answer such questions. This might be a problem particularly, if each one of the resulting rules cover the whole range of the output space, hence the resulting relation hides valuable information. Same danger is applicable to the output clustering approach for the antecedents side. M-A algorithm handles this problem by not projecting the output clusters onto input space one by one, rather keeps the antecedents as NV-dimensional entities and does not forcefully fit a unimodal trapezoidal membership function to the projected clusters.

Since output clustering based approach utilize a single dimensional clustering, a simple grid based clustering algorithm is applicable. This way it can overcome the necessity of determining the "best" level of fuzziness (m) and avoid the anomalies that might be incorporated with the fuzzy clustering algorithms. However, certain strategies must be utilized for the other two approaches in order to decide the most suitable level of fuzziness particularly, if they utilize FCM methodology. Furthermore, single dimensional clustering is always easier to obtain and less open to speculative results since the output space is ordered. For the multidimensional clustering, that's not the case. Usually some sort of prespecified indices must be utilized in order to validate the obtained clusters. And these pre-specified indices themselves are by no means the absolute measure of the validity.

Input clustering algorithms, in general, suffers from the curse of dimensionality. Castellano et al. algorithm handles this problem by just allowing to construct fuzzy if-then rules that corresponds to a fuzzy cluster, i.e. bounds possible number of rules. However it is quite possible that there are more "output clusters" than that one can identify in the *misty* structure of NV-dimensional input space. This would be particularly dangerous for the case when the output is not categorized data.

5   Conclusions

Fuzzy system modelling (FSM) algorithms are prominent data mining tools. There are many different approaches to the structure identification phase of fuzzy modelling. In this paper, we analyzed three possible approaches of incorporating fuzzy clustering in the *structure identification* phase of fuzzy modelling and discussed some issues associated with these algorithms. We also conducted numeric experiments in order to compare these algorithms in terms of their classification accuracy when used in a certain problem of medical diagnosis. Note that, the algorithms used in this study, varies from the other two in other aspects as well, hence the comparison is by no means based on solely the way clustering is utilized. However our experience with the algorithms yields valuable information about the structural problems associated with each approach.

The modified algorithm that is based on the original Sugeno-Yasukawa approach, i.e., output clustering and projecting onto input space, outperforms the other two algorithms in terms of classification performance on the average. The advantage of this algorithm (and the fuzzy clustering approach it utilizes) is the fact that it is the only methodology that keeps the natural ties among the variables through out the all stages of structure identification phase.

The second algorithm proposed by Castellano et al. is based on input clustering approach. The method assumes unimodal convex fuzzy sets, which limits the modelling capability and applicability of the algorithm. One another problem associated with the input clustering approach is the fact that the consequents of the rules might be the same for different rules, i.e. they may coincide with each other. This would be a problem particularly if the consequents span over the whole output range. Such rules deteriorate the descriptive power of the rules and hamper the predictive quality of the model

The third algorithm used in the analysis is proposed by Uncu and Turksen (14), which is based on NV+1 dimensional clustering. Our experience with this approach suggests that a problem source with this algorithm is the fact that NV+1 dimensional clustering treats the output variable not different than an input variable. Hence, again it suffers from the fact that the multidimensional clusters obtained may lead to a rule base where the consequents coincide with each other and result in unrealistic rule bases. Finally, the multidimensional clustering algorithms utilized both in input space clustering approach of Castellano et al. and input-output space clustering algorithm is computationally more cumbersome when compared with a single dimensional clustering algorithm utilized in output clustering approach.

References

[1] H. Axer, J. Jantzen, G. Berks, D. Südfeld, D. v. Keyserlingk, The Aphasia Database on the Web: Description Problems of Classification in Medicine, in: Proceedings of ESIT 2000, Aachen, Germany, 2000, pp. 104-111.

[2] J. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York and London, 1981.

[3] G. Castellano, A. Fanelli, C. Mencar, A Fuzzy Clustering Approach for Mining Diagnostic Rules, in: Proceedings of IEEE Conference on Systems, Man and Cybernetics, Washington, USA, 2003.

[4] G. Castellano, C. Castiello, A. Fanelli, C. Mencar, Knowledge discovery by a neuro-fuzzy modeling framework, Fuzzy Sets and Systems 149 (2005) 187-207

[5] M. Delgado, A. Gomez-Skarmeta, F. Martin, Rapid Prototyping of Fuzzy Models, in: H. Hellendoorn, D. Driankov (Eds.), Fuzzy Model Identification: Selected Approaches, Springer, Berlin, Germany, 1997, pp. 121-161.

[6] K. Kilic, B. Sproule, I. B. Turksen, C. Naranjo, Pharmacokinetic Application of Fuzzy Structure Identification and Reasoning, Information Sciences 162 (2004) 121-137.

[7] M. Mizumoto, Method of Fuzzy Inference Suitable for Fuzzy Control, J. Soc. Instrument and Control Engineers 58 (1989) 959-963.

[8] W. Pedrycz, A. Bargiella, Fuzzy Fractal Dimensions and Fuzzy Modeling, Information Sciences153 (2003) 199-216

[9] W. Pedrycz, K. Keun-Chang, Linguistic models as a framework of user-centric system modeling, IEEE Transactions on Systems, Man and Cybernetics 26 (2006) 727-745

[10] M. Sugeno, T. Yasukawa, A Fuzzy Logic Based Approach to Qualitative Modeling, IEEE Transactions on Fuzzy Systems 1 (1993) 7-31.

[11] T. Takagi, M. Sugeno, Fuzzy Identification of Systems and Its Application to Modelling and Control, IEEE Transactions on Systems, Man and Cybernetics 15 (1985) 116-132.

[12] G. Tsekouras, H. Sarimveis, E. Kavakli, G. Bafas, A hierarchical fuzzy-clustering approach to fuzzy modeling, Fuzzy Sets and Systems 150 (2005) 245-266

[13] O. Uncu, I. B. Turksen, A Novel Feature Selection Approach: Combining Feature Wrappers and Filters, Information Sciences 177 (2007) 449-466.

[14] O. Uncu, I. B. Turksen, K. Kilic, Localm-fsm: A New Fuzzy System Modeling Approach Using a Two-Step Fuzzy Inference Mechanism Based on Local Fuzziness Level, in: Proceedings of International Fuzzy Systems Association World Congress, 2003, pp. 191-194.

[15] J. Yen, Fuzzy Logic - A Modern Perspective, IEEE Transaction on Knowledge and Data Engineering 11 (1999) 153-166.