Mathematical Structure of Fuzzy Modeling of Medical Diagnoses by Using Clustering Models

R.W. W. Hndoosh
Mathematical Structure of Fuzzy Modeling of Medical Diagnoses by Using Clustering Models

Rana Waleed Hndoosh, M. S. Saroa, Sanjeev Kumar

Abstract—An Adaptive-Network-based Fuzzy Inference System ANFIS with different techniques of clustering is successfully developed to solve one of the problems of medical diagnoses, because it has the advantage of powerful modeling ability. In this paper, we propose the generation of an adaptive neuro-Fuzzy Inference System model using different clustering models such as a subtractive fuzzy clustering (SFC) model and a fuzzy c-mean clustering (FCM) model in the Takagi-Sugeno (TS) fuzzy model for selecting the hidden node centers. An experimental result on datasets of medical diagnoses shows the proposed model with two models of clustering (ANFIS-SFC & ANFIS-FCM) while comparing the same model but both with and without clustering models (ANFIS). We obtained better results of average Training error of training and checking data with ANFIS-SFC when we used a Back-propagation model of the Learning Rule, and similarly we obtained the best results with ANFIS-FCM when used with a Hybrid model. Also we have applied SFC & FCM models without ANFIS to get different matrices of cluster centers on medical diagnoses. Finally, we have displayed the surface of MF to each of the ten separate clusters of diseases with values of the objective function.

Index Terms—Clustering, Fuzzy-Cluster System, FCM, SFC, ANFIS, TS Fuzzy Model, Learning Rule.

1. INTRODUCTION

Fuzzy models describe systems by establishing relations between the relevant variables in the form of if-then rules. Traditionally, a fuzzy model is built by using expert knowledge in the form of linguistic rules. Different approaches have been proposed for this purpose, like fuzzy modeling [1], [2], [10], [17] neural-network training techniques [4], [14], [18], [20] and product-space clustering [3], [6], [12]. Solutions to this problem have been sought for fuzzy neural networks [2], [11], [13], [17], and for fuzzy rule-based models in general [10], [16], [18], [20]. In section II, we present a general model of a fuzzy-cluster system that is explained by some steps. Section 3 shows how a mathematical structure of the applied Takagi-Sugeno (TS) rule-based model depends on clustering [3], [9], [16], [19]. In section 4, we describe a fuzzy clustering of TS models that consist of two models, a subtractive fuzzy clustering SFC model and a fuzzy c-mean clustering (FCM) model that are applied to a modeling problem [6], [7], [9], [10], [17], [18]. After explaining steps of SFC and FCM models, in Section 5, we have built the structure of ANFIS with the TS model that depends on FCM and we use a Learning Rule as the "back-propagation learning rule" [8], [5], [11], [13], [18], [20]. In section 6, we apply all these models on real data of medical diagnosis and we perform a system at 50 epochs to get the average testing error of training data and the average testing error of checking of ANFIS-SFC, ANFIS-FCM and ANFIS without clustering in part of the experiments and discussion [9], [10], [15]. The performance of the models is measured by the square root of the mean squares prediction error (RMSE), where we obtained very small errors, and the smallest value of Average RMSE with ANFIS-FCM. So we have built a system that has proved its efficiency in the medical diagnoses. Finally, some concluding remarks are given in section 7.

2. GENERAL FORMULATION OF A FUZZY-CLUSTER SYSTEM

Fuzzy clustering is used to detect multidimensional fuzzy sets in the product space of the input variables to identify the antecedent of the fuzzy rules and then assign a linear consequent to each rule. The fuzzy rule base comprises the following fuzzy IF-THEN rules [10], [17]:

\[ R^{(i)}: IF \ x_1 \ is \ A_1^i \ and \ ... \ and \ x_n \ is \ A_n^i \ THEN \ \hat{y} \ is \ B^i, \]

where \( A_1^i \) and \( B^i \) are fuzzy sets in \( U_i \subset R, V \subset R \), respectively, and \( x = (x_1, x_2, ..., x_n) \in U \) and \( y \in V \) are the input and output (linguistic) variables of the fuzzy system, respectively [4], [7], [11].

Suppose that we are given \( C \) input-output pairs \( (x_0^l, y_0^l), \ l = 1, 2, ..., C \). Our task is to construct a fuzzy system \( \hat{y} \) that can match all the \( C \) pairs to any given accuracy; that is, for any given \( \epsilon > 0 \), we require that \( |y_0^l - \hat{y}_0^l| < \epsilon, \ \forall \ l = 1, 2, ..., C \).

This optimal fuzzy system may be constructed as [18]:

\[ \hat{y} = \frac{\sum_{i=1}^{C} y_0^l \cdot \mu_i(x)}{\sum_{i=1}^{C} \mu_i(x)} \]  

Clearly, the fuzzy system (2) is constructed from the \( C \) rules in the form of (0) with \( \exp(-\frac{x_1-x_0^l}{\sigma^2}) \) and the center of \( B^i \) is equal to \( y_0^l \) and uses the product inference system, singleton fuzzifier, and center average defuzzifier. The optimal fuzzy...
system (2) uses one rule for one input-output pair. Thus it is no longer a practical system if the number of input-output pairs is large [12]. For these large sample problems, various clustering techniques can be used to group the input-output pairs so that a group can be represented by one rule. In this work, we have used [2]:

**Step1:** Starting with the first input-output pair \((x_0^1, y_0^1)\), establish a cluster center \(x_0^1\) at \(x_0^1\), and set \(A^1(1) = y_0^1, B^1(1) = 1\). Select a radius \(r\).

**Step2:** Suppose that when we consider the \(k\)th input-output pair \((x_k^l, y_k^l)\), \(k = 2, 3, \ldots\), there are \(C\) clusters with centers at \(x_c^1, x_c^2, \ldots, x_c^C\). Compute the distances of \(x_0^1\) to these \(C\) cluster centers, \(|x_0^1 - x_c^l|, l = 1, 2, \ldots, C\), and let the smallest distances be \(|x_0^1 - x_c^l|\), that is, the nearest cluster to \(x_0^1\) is \(x_c^l\) then:

a) If \(|x_0^1 - x_c^l| > r\), establish \(x_0^l\) as a new cluster center \(x_c^{l+1} = x_0^l\), set \(A^{l+1}(k) = y_k^l, B^{l+1}(k) = 1\), and keep \(A^l(k) = A^l(k - 1), B^l(k) = B^l(k - 1)\), for \(l = 1, 2, \ldots, C\).

b) If \(|x_0^1 - x_c^l| \leq r\), do the following:

\[
A^l(k) = A^l(k - 1) + y_k^l, \tag{3}
\]

\[
B^l(k) = B^l(k - 1) + 1, \tag{4}
\]

\[
A^l(k) = A^l(k - 1), \tag{5}
\]

\[
B^l(k) = B^l(k - 1), \tag{6}
\]

**Step3:** If \(x_0^l\) does not establish a new cluster then the designed fuzzy system based on the \(k\) input-output pairs \((x_j^l, y_j^l), j = 1, 2, \ldots, k\), is:

\[
\hat{y} = \frac{\sum_{l=1}^{C} A^l(k) e^{-\frac{|x_j^l - x_c^l|^2}{\sigma^2}}}{\sum_{l=1}^{C} B^l(k) e^{-\frac{|x_j^l - x_c^l|^2}{\sigma^2}}} \tag{7}
\]

If \(x_0^l\) establishes a new cluster, then the designed fuzzy system is [3], [7]:

\[
\hat{y} = \frac{\sum_{l=1}^{C} A^l(k) e^{-\frac{|x_j^l - x_c^l|^2}{\sigma^2}}}{\sum_{l=1}^{C} B^l(k) e^{-\frac{|x_j^l - x_c^l|^2}{\sigma^2}}} \tag{8}
\]

**Step4:** Repeat by going to (Step2) with \(k = k + 1\).

From (3) to (6) we see that the variable \(B^l(k)\) equals the number of input-output pairs in the \(l\)th cluster after \(k\) input-output pairs have been used, and \(A^l(k)\) equals the summation of the output values of the input-output pairs in the \(l\)th cluster. Therefore, if each input-output pair establishes a cluster center, then the designed fuzzy system (8) becomes the optimal fuzzy system (2), because the optimal fuzzy system (2) can be viewed as using one rule to match one input-output pair and the fuzzy system (7) or (8) can be viewed as using one rule to match one cluster of input-output pairs. The number of clusters (or rules) depends on the distribution of the input points in the input-output pairs and the radius \(r\).

The radius \(r\) determines the complexity of the designed fuzzy system. For smaller \(r\), we have more clusters. For larger \(r\), the designed fuzzy system is simpler but less powerful. In practice, a good radius \(r\) may be obtained by test and errors [4], [10], [18], [20].

**3. MATHEMATICAL STRUCTURE OF THE APPLIED TS MODELS**

The unknown nonlinear system of the TS fuzzy model is formulated as: \(\hat{y}_k = f(x_k)\), that is based on specified or measured input data \(x_k = [x_{1k}, x_{2k}, \ldots, x_{ck}]^T\) and measured output data \(y_k\) of the system, where \(k = 1, 2, \ldots, N\) denotes the index of the \(k\)th input-output data-pair. In general it may not be easy to find a global nonlinear model that is universally applicable to describe the unknown system \(f(.)\). In that case it would certainly be to build local linear models for specific operating points of the process and combine these into a global model. This can be done by combining a number of local models, where each local model has a predefined operating region. This results in the so-called operating of the TS fuzzy model when the operating region is defined by fuzzy rules. This type of operating is formulated as [3], [9], [14]:

\[
\hat{y}_k = \sum_{l=1}^{C} w_l(x_k) (\alpha^T \phi_k + b_l), \tag{9}
\]

where \(w_l(x_k)\) describes the operating of the \(l\)th local linear model defined by the parameter vector \(\alpha_l = [a_l^T b_l]^T\), and \(x_k, \phi_k\) subsets of the original input vector \(x_k\). From the general form (0) with the form of the TS fuzzy model we derive the model as:

\[
R(l): IF \, x_{1k} \, is \, A_1^l(x_{1k}) \, and \, \ldots \, \text{and} \, x_{nk} \, is \, A_n^l(x_{nk}) \, THEN \, \hat{y}_k = a_l^T \phi_k + b_l \, [\omega_l] \tag{10}
\]

\(A^l_j(x_{jk})\) represents a multivariate membership function that describes the fuzzy set \(A^l_j\). Where \(\alpha_l\) and \(b_l\) are the parameters of the local linear model, and \(\omega_l = [0, 1]\) is the weight of the rule that represents the desired impact of the rule. (It is often chosen by the designer of the fuzzy system). The degree of performance of a rule is calculated as the product of the degree of performance of the fuzzy sets in the rule that is denoted by [11], [13], [19]:

\[
\beta_l(x_k) = \omega_l \beta_l^1(x_k) = \omega_l \prod_{j=1}^{n} A_j^l(x_{jk}) \tag{11}
\]

The rules of the fuzzy model are aggregated using the normalized fuzzy mean formula:

\[
\hat{y}_k = \frac{\sum_{l=1}^{C} \omega_l \beta_l(x_k) (\alpha_l^T \phi_k + b_l)}{\sum_{l=1}^{C} \omega_l \beta_l(x_k)} \tag{12}
\]

Gaussian membership functions are used here to represent the fuzzy set \(A^l_j(x_{jk})\):

\[
A^l_j(x_{jk}) = \exp \left(-\frac{(x_{jk}-c_{jk})^2}{\sigma^2}ight), \tag{13}
\]
where $c_{ij}$ represents the center and $\delta_{ij}^2$ the variance of the Gaussian curve. The use of Gaussian membership function allows for the compact formulation of (11), [4]:

$$\beta_i(x_k) = \omega_i A_i(x_k) = \omega_i \exp \left( -\frac{1}{2} (x_k - c_j)^T (F_i)^{-1} (x_k - c_j) \right).$$

(14)

where $c_j = [c_{j1}, c_{j2}, \ldots, c_{jc}]$ denotes the center of the $l^{th}$ multivariate Gaussian and $F_i$ is the inverse of the matrix containing the variances on its diagonal:

$$F_i = \begin{pmatrix}
\delta_{11}^2 & 0 & \cdots & 0 \\
0 & \delta_{22}^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \delta_{cc}^2
\end{pmatrix}$$

(15)

In the following sections a special type of clustering model is presented.

4. FUZZY CLUSTERING of TS MODELS

In this section, we will review the concepts of two different models, the first is subtractive fuzzy clustering model and the second is Fuzzy c-means model.

4.1 Subtractive Fuzzy Clustering (SFC) Model

Subtractive fuzzy clustering divides the $X$ data into $c$ fuzzy clusters. This means, each observation consists of the input data into $X$ data into $Subtractive fuzzy clustering divides the clusters. This means, each observation consists of the input data into $X$ data into

In the following sections a special type of clustering model is presented.

4. FUZZY CLUSTERING of TS MODELS

In this section, we will review the concepts of two different models, the first is subtractive fuzzy clustering model and the second is Fuzzy c-means model.

4.1 Subtractive Fuzzy Clustering (SFC) Model

Subtractive fuzzy clustering divides the $X$ data into $c$ fuzzy clusters. This means, each observation consists of the input data into $X$ data into $Subtractive fuzzy clustering divides the clusters. This means, each observation consists of the input data into $X$ data into
where \( \mu_{ik} \) are the elements of the matrix and represent the degree of the membership. The fuzzy c-means functional \( J_q \) is called objective function defined as [10], [12], [13], [18]:

\[
J_q(U, c) = \sum_{k=1}^{n} \sum_{l=1}^{C} \mu_{lk}^q d^2(x_k, c_l),
\]

(26)

where \( U \in \mu_{fc} \) is a fuzzy c partition of \( x \); \( c = (c_1, c_2, ..., c_C) \) the clusters, \( 1 \leq l \leq C \), and \( d_{kl} = \|x_k - c_l\|^2 \). The basic steps of the model repeat for \( q \in [1, \infty) \) (\( q \) is the iteration counter) that are given as follows [7]:

Step1: Calculate the parameters of the clusters:

1. Calculate the centers and standard deviation of the Gaussian membership functions:

\[
e_{c_1} = \frac{1}{\sqrt{2\pi} \delta_{c_1}}, \quad \delta_{c_1}^2 = \frac{1}{\sum_{k=1}^{n} \mu_{lk}^{q-1}}.
\]

(27)

2. Calculate the parameters of the clusters:

\[
e_{c_k} = \frac{1}{\sqrt{2\pi} \delta_{c_k}}, \quad \delta_{c_k}^2 = \frac{1}{\sum_{l=1}^{C} \mu_{lk}^{q-1}}.
\]

(28)

3. Calculate the parameters of the local models:

4. We can calculate \( \theta_l \) by the same procedure as in the SFC model:

\[
\theta_l = (\Phi_l^T \beta_l \phi_k)^{-1} \Phi_l^T \beta_l y.
\]

(30)

5. Calculate weights of the rules:

\[
\omega_l = \frac{1}{\left( \sum_{k=1}^{n} \mu_{lk} \right)^{\frac{1}{q-1}}}.
\]

(31)

Step2: Compute the distance measure \( d_{jk}^2 \):

The distance measure consists of two terms. The first term is based on the geometrical distance between the \( c_k \) cluster centers and \( x_j \) input vector, while the second is based on the performance of the linear models:

\[
d_{jk}^2 = \frac{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}}{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}} \left( \frac{\sum_{l=1}^{n} \mu_{lk} \mu_{kl}}{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}} \right)
\]

(32)

\[
= \frac{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}}{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}} \left( \frac{\sum_{l=1}^{n} \mu_{lk} \mu_{kl}}{\prod_{l=1}^{n} \mu_{lk} \mu_{kl}} \right)
\]

(33)

Step3: Update the partition matrix:

\[
\mu_{ik}^{(m+1)} = \frac{1}{\sum_{l=1}^{C} \mu_{lk}^{m}}, \quad 1 \leq l \leq C, 1 \leq k \leq n.
\]

(34)

until \( \|U^{(m)} - U^{(m-1)}\| < \varepsilon \), stop; otherwise.

5. STRUCTURE OF ANFIS USING CLUSTERING with TS MODEL

Suppose \( x = [x_1, x_2, ..., x_n] \) is the input vector, \( c_k = [c_{1k}, c_{2k}, ..., c_{nk}] \) is the center cluster of membership functions in the \( k^{th} \) neuron, and \( \delta_k = [\delta_{1k}, \delta_{2k}, ..., \delta_{nk}] \) is the width vector of membership functions in the \( k^{th} \) neuron, [8], [14].

1. The Input layer is the first layer in ANFIS, where each neuron in this layer represents an input variable \( x = (x_1, x_2, ..., x_n)^T \):

\[
\hat{y}_1 = \mu_{A_k}(x_k) = \exp \left( \frac{1}{2\sigma_k^2} \right)
\]

(35)

2. The Hidden layer consists of a fuzzifier layer, a normalized layer, and a weighted layer. In this layer all the steps that are explained above for SFC and FCM of TS model are applied:

\[
\hat{y}_2 = \exp \left( \frac{\sum_{k=1}^{n} \left( x_k - c_{lk} \right)^2}{2\sigma_k^2} \right)
\]

(36)

\[
\hat{y}_3 = \exp \left( \frac{\sum_{k=1}^{n} \left( x_k - c_{lk} \right)^2}{2\sigma_k^2} \right)
\]

(37)

\[
\hat{y}_4 = \hat{y}_3 \cdot \left( a_i^T \phi_k + b_i \right)
\]

(38)

3. The Output layer. The output of the neuron in this layer is:

\[
\hat{y}_5 = \sum_{i=1}^{n} \hat{y}_4
\]

(39)

where \( \hat{y} \) is the value of an output variable.

A feed forward adaptive network is actually a static mapping between its input and output spaces; this mapping may be either a simple linear relationship or a highly nonlinear one, depending on the structure for the network and the function for each node [2], [19], [20]. Here our aim is to construct a network for achieving a desired nonlinear mapping that is regulated by a data set consisting of a number of desired input-output pairs of a target system. This data set is usually called the training data set and the procedure we follow in adjusting the parameters to improve the performance of the network is often referred to as the learning rule. Usually an adaptive network’s performance is measured as the discrepancy between the desired output and the network’s output under the same input conditions. This discrepancy is called the error measure.

5.1 Mathematical Model of the Learning Rule

The central part of a learning rule for an adaptive network is concerned with how to recursively get a slope vector in which each element is defined as the derivative of an error measure with respect to a parameter [9], [11], [13]. The model is
To calculate the derivative of the overall error measure $E$ with respect to $\beta$, we have:

$$\frac{\delta E}{\delta \beta} = \sum_{p=1}^{P} \frac{\delta E_p}{\delta \beta}$$

where $k$ is the step size, the length of each slope transition in the parameter space.

### Table 1. The inputs with range selected respectively.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Symptoms (Symbol)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>Fever (F.)</td>
<td>[99-107]</td>
</tr>
<tr>
<td>$X_2$</td>
<td>Malaise (ma.)</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_3$</td>
<td>Anorexia (An.)</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_4$</td>
<td>Headache (He.)</td>
<td>[0.25-0.75]</td>
</tr>
<tr>
<td>$X_5$</td>
<td>Cough (Co.)</td>
<td>[0.05-0.95]</td>
</tr>
<tr>
<td>$X_6$</td>
<td>Abdominal Pain (Ab.)</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_7$</td>
<td>Diarrhea (D.)</td>
<td>(5-8) times P. 12 h.</td>
</tr>
<tr>
<td>$X_8$</td>
<td>Lethargy (Le.)</td>
<td>[0.05-0.4]</td>
</tr>
<tr>
<td>$X_9$</td>
<td>Skin Rash (SR.)</td>
<td>[0.10-0.80]</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>Myalgia (My.)</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{11}$</td>
<td>Delirium (De.)</td>
<td>[0.20-0.80]</td>
</tr>
<tr>
<td>$X_{12}$</td>
<td>Difficulty concentrating Dc.</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{13}$</td>
<td>Mild coryza (Mcr.)</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{14}$</td>
<td>Enlarged lymph &amp; pain (Ep.)</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_{15}$</td>
<td>Moderate Throat pain (Mp.)</td>
<td>[0.10-0.80]</td>
</tr>
<tr>
<td>$X_{16}$</td>
<td>Rhinitis (Rh.)</td>
<td>[0.20-1]</td>
</tr>
<tr>
<td>$X_{17}$</td>
<td>Hoarseness of voice (Hv.)</td>
<td>[0.25-0.85]</td>
</tr>
<tr>
<td>$X_{18}$</td>
<td>Rigor (Ri.)</td>
<td>(2-10) times P. 12 h.</td>
</tr>
<tr>
<td>$X_{19}$</td>
<td>Swelling of salivary glands Ss.</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_{20}$</td>
<td>Difficulty in Swallowing DS.</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{21}$</td>
<td>Skin Rash with pruritus SR.</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_{22}$</td>
<td>Vomiting (v.)</td>
<td>(2-10) times P. 12 h.</td>
</tr>
<tr>
<td>$X_{23}$</td>
<td>Difficulty breathing (Db.)</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{24}$</td>
<td>Paleness (pl.)</td>
<td>[0.20-0.80]</td>
</tr>
<tr>
<td>$X_{25}$</td>
<td>Muscle cramp (Mcr.)</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{26}$</td>
<td>Sickness (S.)</td>
<td>[0.25-0.75]</td>
</tr>
<tr>
<td>$X_{27}$</td>
<td>Heart beat (Hb.)</td>
<td>(85-250) time P. m.</td>
</tr>
<tr>
<td>$X_{28}$</td>
<td>Paralysis (pr.)</td>
<td>[0.10-0.70]</td>
</tr>
<tr>
<td>$X_{29}$</td>
<td>Thyroid swelling (Ts.)</td>
<td>[0.10-0.90]</td>
</tr>
<tr>
<td>$X_{30}$</td>
<td>Pain in Thyroid gland (Pt.)</td>
<td>[0.10-0.80]</td>
</tr>
<tr>
<td>$X_{31}$</td>
<td>Blood with sp &amp; phlegm Bsp.</td>
<td>[0-1]</td>
</tr>
<tr>
<td>$X_{32}$</td>
<td>Night sweats (Ns.)</td>
<td>[0.25-0.75]</td>
</tr>
</tbody>
</table>

6. The PROPOSED MODEL

In this section, we will apply all previous concepts on application of medical diagnosis and the comparison the results of $SFC$ and $FCM$ without using ANFIS as following:

6.1 Description of Problem

In this application of medical diagnosis, we chose acceptable and real data because the system has to deal with the patient’s health and even life, and this real data depends on a continuous range of numbers according to a range of symptoms [15]. The data base has given us a group of doctors...
and staff specialists from a combination of 3 hospitals (“Mosul General Hospital”, “Al Kindi General Teaching Hospital”, and “Specialized Hospital Eben Al Athir for Children”) merged into a single dataset for research purposes. Table (1) illustrates the symptoms (inputs from $x_1$ to $x_{32}$) selected respectively with a range for all symptoms.

We selected 100 patients from the information of a real dataset for many visitors of a hospital with ten patients per disease for a total of 10 different diseases (Typhoid Fever, Rubella, Acute Laryngitis, Mumps, Chicken pox, Malaria, Anemia, Polio, Goiter, and Tuberculosis). Table (2) shows the list of diseases and their assigned labels from 1 to 10 respectively. This dataset provided 32 symptoms of ten different diseases. We have processed the real data to conduct the process of scaling to all data based on linear equations and obtained scaled data limited between 0 and 1. From these scaled data we obtained a matrix of dimensions (100x32) for a hundred patients (in rows), where the 32 column indicate the symptoms (inputs) for ten different diseases (output). Each disease depends on some of the symptoms: that is the relationship between inputs and output. The data matrix has been identified depending on this relationship between symptoms and disease that correlate with these symptoms as shown in table (3). Representing the real data we have available for the application of medical diagnostics enables us to draw the clusters in separate groups and in points. Each group has a different color, so that each disease is represented by a different color. See Fig. (1).

6.2 Experiments and Discussion

We will now load the inputs and output variables into the workspace. Data inputs have 32 columns (symptoms) and data output has 1 column (diseases) representing the output variable, as shown in Fig.(2). The number of rows is 100 and represents the number of patients. The columns from 1 to 32 represent the values of variables for the symptoms, and the last column (33rd column) represents the output values of the variable of the diseases. We modeled the relationship between the input and the output variables by using the concept of clustering of the data.

<table>
<thead>
<tr>
<th>Disease name (output)</th>
<th>Assigned label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Typhoid Fever</td>
<td>6</td>
</tr>
<tr>
<td>2 Rubella</td>
<td>7</td>
</tr>
<tr>
<td>3 Acute Laryngitis</td>
<td>8</td>
</tr>
<tr>
<td>4 Mumps</td>
<td>9</td>
</tr>
<tr>
<td>5 Chicken pox</td>
<td>10</td>
</tr>
<tr>
<td>6 Malaria</td>
<td>1</td>
</tr>
<tr>
<td>7 Anemia</td>
<td>2</td>
</tr>
<tr>
<td>8 Polio</td>
<td>3</td>
</tr>
<tr>
<td>9 Goiter</td>
<td>4</td>
</tr>
<tr>
<td>10 Tuberculosis</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3. The relationship between inputs and output.

Now we will build an adaptive neuro-fuzzy inference system (ANFIS) based on the clustering process, and the first thing we do is to load data in the workspace. Of the original 100 patients, we use 75 patients as training data and 25 patients as checking data. After loading data, we generate a model from data using clustering, to determine which method we will use in the process of clustering. We can generate an FIS structure using subtractive fuzzy clustering (SFC) and fuzzy c-main (FCM) models. To use the clustering technique, we should select the clustering type of the ANFIS program before the FIS is generated.

The SFC and FCM models partition the data into groups called clusters, and generate an FIS with the minimum number of rules required to distinguish the fuzzy qualities associated with each of the clusters. The FIS is composed of inputs, output and rules. See Fig.(3). The number of rules equals the number of clusters and hence 10 rules are created.
Notice, all the inputs and outputs have exactly 10 membership functions for each input variable. The 10 membership functions represent the 10 clusters that were identified by SFC & FCM. Notice that the membership function type is “gaussmf” (Gaussian MF) and that the parameters of the membership function are linear, that is denoted by the mathematical formula:

$$\mu(x;a,b) = \exp\left(-\frac{(x-a)^2}{2b^2}\right),$$

where $a, b$ parameters

$$R_i : \text{IF } x_1 \text{ is } A_1^{(i)} \text{ and } x_2 \text{ is } A_2^{(i)} \text{ and } \ldots \text{ and } x_{32} \text{ is } A_{32}^{(i)} \text{ THEN }$$

$$f_1 = a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,32}x_{32} + a_{1,33}$$

$$R_{i,j} : \text{IF } x_1 \text{ is } A_1^{(i)} \text{ and } x_2 \text{ is } A_2^{(j)} \text{ and } \ldots \text{ and } x_{32} \text{ is } A_{32}^{(j)} \text{ THEN }$$

$$f_{ij} = a_{ij,1}x_1 + a_{ij,2}x_2 + \ldots + a_{ij,32}x_{32} + a_{ij,33}$$

where $x_1, x_2, \ldots, x_{32}$ are the symptoms variables. See table (1). $A_j^{(i)}$ MF is according to variables ($i = 1 \text{ to } 10$ and $j = 1 \text{ to } 32$), $\gamma_-$ are the rules, $\gamma_+$ as the linear function of output (diseases), $\{\gamma_{i,j}\}$ are the parameters. Let’s explore how the fuzzy rules are constructed. There are exactly 10 rules. Each rule attempts to map a cluster in the input space to a cluster in the output space. The first rule can be explained simply as follows: If the inputs to the FIS are High Fever, Malaise, Anorexia, Headache, Cough, Abdominal Pain, Diarrhea, Lethargy, Skin Rash, Myalgia, Delirium, and Difficulty concentrating (all of which strongly belong to their respective cluster1 MFs) then the output “Typhoid Fever” must strongly belong to its cluster1 MF. The other rules in the output space the function similarly.

The output of the FIS, diseases, has 10 linear MFs (ST model) representing the 10 clusters identified by the SFC & FCM models. Parameters are estimated from the dataset using the least squares estimation technique by (24), (30). We can now use the FIS that has been constructed by the ST model with the SFC & FCM models to understand the relationship being modeled. We can use command windows, “surfview” the surface viewer that helps view the input-output surface of the fuzzy system. This conception is very helpful to understand how the system is going to behave for the entire range of values in the input space, as shown in Fig. (4).

Now, having built the fuzzy system, if we want to understand which disease will occur for particular symptoms, than this tool will help us to simulate the FIS response for the input of our choice. After completing the system, we save it, and will go back again to the ANFIS and load FIS from file. Now, an FIS structure is returned; the model type for the FIS structure is a first order TS model with 10 rules that are applied the first time with the SFC model and the second time with the FCM model. From the program of ANFIS the structure of the system appears as shown in Fig.(5). We have selected a type of the Learning Rule (back-propagation learning rule) as an ANFIS parameter optimization method for FIS training. This reduces the problem of too much propagation of rules when the input data has a high dimension (as in our application). In the next step, we perform a system at 50 epochs and get the average testing error of training data and the average testing error of checking of ANFIS-SFC, ANFIS-FCM and ANFIS without clustering. Additionally, the number of nodes are a number of linear parameters and a number of nonlinear parameters, as shown in table(4) is obtained where the better results of Training error, ATE of training data and ATE of checking data was obtained with ANFIS-SFC (0.013637) when the Back-propagation model of the Learning Rule was used. The best results of Training error, ATE of training data and ATE of checking data was obtained with ANFIS-FCM (0.0002172) when the Hybrid model of the Learning Rule was used. See Fig.(6a, 6b, 6c). We tried to compare the real output of medical diagnoses with the results that we got from the systems for three cases namely with ANFIS-SFC, ANFIS-FCM and ANFIS without clustering used. The performance of the models is measured by the square root of the mean squares prediction error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

(51)

Using (51) we obtained a very small error, and the smallest value of Average RMSE was obtained with ANFIS-FCM, as the following table (5) illustrates. So we have built a system that has proved its efficiency in the medical diagnoses.

In the second part of our application, we have applied the clustering without a neural network on the data.
that is represented by 100 cases. The first time, we applied the FCM model depending on (26) to (34) with 10 clusters. The second time, we applied the SFC model depending on (16) to (25) with the same number of clusters. When these steps were applied to the SFC model by using MATLAB, we obtained the matrix of optimal cluster centers $C$. See Fig.(7a). Each row of $C$ contains the position of a cluster center.

We also obtained the vector $S$ that contains the sigma values that specify the range of effect of a cluster center in each of the data dimensions. All cluster centers share the same set of sigma values. We represented the cluster centers with clusters as in Fig.(8), where the symbols ‘o’ represent data points for each cluster component and a big circle in the center of each cluster represents the cluster center. Each cluster is represented by a different color. However, when these steps are applied to the FCM model, we have obtained a matrix of final cluster centers where each row provides the center coordinates. See Fig.(7b). We also obtained final membership function matrix $U$. The plot in Fig.(9) displays the surface of $MF$ to each of the ten separate clusters, and values of the objective function during iterations ($obj-fcn$). This value started from 132.22 and went up to 10.20 at iteration 22, as shown in Fig.(10).

Finally, when we have selected the number of clusters ($N=10$), we got to 10 clusters after applying clustering programs, where each cluster includes 10 patients suffering from the same disease, where:

The First cluster contains all patients who have symptoms of Typhoid Fever.
The Second cluster contains all patients who have symptoms of Mumps.
The Third cluster contains all patients who have symptoms of Acute Laryngitis.
The Fourth cluster contains all patients who have symptoms of Malaria.

The Fifth cluster contains all patients who have symptoms of Chicken pox.
The Sixth cluster contains all patients who have symptoms of Goiter.
The Seventh cluster contains all patients who have symptoms of Rubella.
The Eighth cluster contains all patients who have symptoms of Polio.
The Ninth cluster contains all patients who have symptoms of Anemia.
The Tenth cluster contains all patients who have symptoms of Tuberculosis.

<table>
<thead>
<tr>
<th>Type of Learning Rule</th>
<th>Type modeling</th>
<th>Training error</th>
<th>ATE of training data</th>
<th>ATE of checking data</th>
<th>No. of nodes</th>
<th>No. of L.Ps</th>
<th>No. of non-L.Ps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back-propagation</td>
<td>ANFIS-SFC</td>
<td>0.013637</td>
<td>0.01363</td>
<td>0.0041308</td>
<td>695</td>
<td>330</td>
<td>640</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
<td>0.00015227</td>
<td>0.00014509</td>
<td>0.13004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back-propagation</td>
<td>ANFIS-FCM</td>
<td>0.321</td>
<td>0.32017</td>
<td>0.35873</td>
<td>695</td>
<td>330</td>
<td>640</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
<td>0.00002172</td>
<td>0.00002174</td>
<td>0.07341</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Back-propagation</td>
<td>ANFIS</td>
<td>0.2341</td>
<td>0.23271</td>
<td>0.22721</td>
<td>119</td>
<td>10</td>
<td>96</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
<td>0.14078</td>
<td>0.14078</td>
<td>0.19845</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 5. Comparison of the results of ANFISs with real outputs. |
|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|</p>
<table>
<thead>
<tr>
<th>No</th>
<th>Real output</th>
<th>ANFIS output</th>
<th>Error RMSE</th>
<th>ANFIS-SFC output</th>
<th>Error RMSE</th>
<th>ANFIS-FCM output</th>
<th>Error RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>6.8</td>
<td>0.0365</td>
<td>7</td>
<td>0</td>
<td>7.01</td>
<td>0.0018</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>6.99</td>
<td>0.0018</td>
<td>6.9</td>
<td>0.0183</td>
<td>6.92</td>
<td>0.0146</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>6.8</td>
<td>0.0365</td>
<td>6.74</td>
<td>0.0475</td>
<td>6.81</td>
<td>0.0347</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.99</td>
<td>0.0018</td>
<td>0.933</td>
<td>0.0122</td>
<td>0.99</td>
<td>0.0018</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.98</td>
<td>0.0037</td>
<td>1.05</td>
<td>0.0091</td>
<td>0.96</td>
<td>0.0073</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0.99</td>
<td>0.0018</td>
<td>0.826</td>
<td>0.0318</td>
<td>0.95</td>
<td>0.0091</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3.2</td>
<td>0.0365</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2.94</td>
<td>0.0110</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3.23</td>
<td>0.0420</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3.99</td>
<td>0.0018</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3.97</td>
<td>0.0055</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>3.99</td>
<td>0.0018</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>10.02</td>
<td>0.0037</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>9.93</td>
<td>0.0128</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

http://www.ijser.org
systems have been proven to produce better results compared to other techniques. In this work we compared three different techniques, first \textit{ANFIS-SFC}, second \textit{ANFIS-FCM} and third \textit{ANFIS} without clustering, where each one of these techniques depends on different models of clustering. From that we noted the best results of the \textit{Training error}, the \textit{ATE} of the training data and the \textit{ATE} of checking data with \textit{ANFIS-SF} (0.013637) when the \textit{Back-propagation} model of the Learning Rule was used. We have obtained the best results of the \textit{Training error}, the \textit{ATE} of the training data and the \textit{ATE} of the checking data with \textit{ANFIS-FCM} (0.00002172) when the Hybrid model of the Learning Rule was used. The smallest value of the average \textit{RMSE} was obtained with the \textit{ANFIS-FCM} (0.0056). Also we have compared the results of \textit{SFC} and \textit{FCM} without using \textit{ANFIS}, where results for different matrices of cluster centers are denoted by fig.(9a) and (9b). However, in both models we got 10 \textit{clusters} and each cluster includes 10 patients suffering from the same disease. We note that these techniques are very effective with mathematical modeling of medical diagnosis. We can improve this work by using other models of clustering like G-K clustering, that depends on the concept of fuzzy modeling or improves this system to find treatment of each case.

ACKNOWLEDGEMENTS

The author wishes to thank the reviewers for their excellent suggestions that have been incorporated into this paper. I would like to thank and acknowledge Prof. Deepak Gupta, Head department of mathematics, M.M. University, Mullana, Ambala, India, who advised me to improve my work.

REFERENCE


