Improving Classifications of Medical Data Based on Fuzzy ART2 Decision Trees

Yo-Ping Huang, Shin-Liang Lai, Frode Eika Sandnes, and Shen-Ing Liu

Abstract

Analyzing given medical databases provide valuable references for classifying other patients symptoms. This study presents a strategy for discovering fuzzy decision trees from medical databases, in particular Harbeman’s Survival database and the Blood Transfusion Service Center database. Harbeman’s Survival database helps doctors treat and diagnose a group of patients who show similar past medical symptoms and the Blood Transfusion Service Center database advises individuals about when to donate blood. The proposed data mining procedure involves neural network based clustering using Adaptive Resonance Theory 2 (ART2), and the extraction of fuzzy decision trees for each homogeneous cluster of data records using fuzzy set theory. Besides, another objective of this paper is to examine the effect of the number of membership functions on building decision trees. Experiments confirm that the number of erroneously clustered patterns is significantly reduced compared to other methods without preprocessing data using ART2.

Keywords: Data classification, fuzzy ART2 algorithm, fuzzy decision tree, medical data.

1. Introduction

Data mining, or knowledge discovery from data (KDD), is the process of extracting desirable knowledge, or interesting patterns, from existing databases for specific purposes. Many types of knowledge and technology have been proposed, and the extraction of decision trees from transaction data is one of the most commonly studied forms of data mining.

Decision tree based classification is a supervised learning method that constructs decision tree from a set of samples [1, 2]. A decision tree is a tree structure where each leaf node is assigned a class label. The root node in the tree contains all training samples that are to be divided into classes. All nodes except the leaves are called decision nodes, since they specify decision to be performed at this node based on a single feature. Each decision node has a number of child nodes equal to the number of values that a given feature assumes. All decision tree algorithms are based on Hunt’s concept learning algorithm [3, 4]. The concept learning algorithm mimics how humans learn simple concepts, that is, how key distinguishing features between two categories are found, represented by positive and negative (training) examples. Hunt’s algorithm is based on a divide-and-conquer strategy where the task is to divide the set \( S \), consisting of \( n \) samples belonging to \( c \) classes, into disjoint subsets that create a partition of the data into subsets containing samples from one class only. Decision trees have many advantages, such as rapid training (tree construction) and high classification accuracy.

Artificial neural networks (ANNs) are commonly used for solving pattern classification problems and for finding decision trees [5]. In the medical domain, neural networks and other strategies have been used for diagnostic decision support such as cancer diagnosis [6] and fall detection [31]. Other fault detection models based on abdicative network model, and combined fuzzy logic and neural networks have been proposed [7-11] as well as a hybrid model combining Adaptive Resonance Theory (ART) and fuzzy c-mean clustering for medical classification and diagnosis with missing features [12].

This study proposes a method using fuzzy ART2 to improve data classification accuracy when discovering fuzzy decision trees. The input patterns are first fuzzified. These fuzzified patterns are clustered into groups where patterns have similar properties by the ART2 neural network. Finally, these groups are organized into a decision tree. The erroneous classification rate is greatly reduced and the efficiency of the algorithm in finding decision tree rules is increased since data patterns are clus-
tered in advance. ART2 is computationally effective and allows the user to easily control the degree of pattern similarity within each cluster [30].

The proposed strategy has been applied on, but not limited to, two different medical databases. The first database includes information about patients who have undergone surgery for breast cancer. Those with similar symptoms are used to discover group decision trees rules to assist the doctors in treatment and diagnosis. The second database contains data about blood donors. The purpose is to find decision tree rules that can be used for providing advice about when to donate blood.

2. Related Work

A. Fuzzy Neural Networks

ANN learning algorithms are either supervised or unsupervised. Supervised learning involves training instances, or examples, with labeled outputs. Unsupervised learning involves an unknown goal with no pre-determined categorizations. ANN technology helps summarize, organize, and classify data. Requiring a few assumptions, they also identify patterns among input data with a high predictive accuracy [13].

Both ANNs and fuzzy models have been successfully applied in many areas [9-11, 14-16] and approaches for successfully combining these two approaches have become an active area of research. A fuzzy neural network (FNN) system uses the ANN learning algorithm to produce parameters. It then adapts these parameters for optimization.

B. Fuzzy ID3 Decision Trees

A decision tree is a simple structure, built through supervised learning, where nonterminal nodes represent tests on one or more attributes and terminal nodes reflect decision outcomes. Decision trees have been successfully applied to classification problems and several decision tree construction algorithms have been proposed [17-19, 34-36]. Decision tree induction methods such as ID3, C4.5 and CART [20-22] generated a tree structure through recursively partitioning the attribute space until the whole decision space is completely partitioned into a set of non-overlapping subspaces. Besides, through the pruning processes, decision trees can be further simplified to classify data without sacrificing their accuracy [34].

Fig. 1 shows an example of a decision tree obtained using the C4.5 algorithm [2] created from Roiger and Geats data [23]. The decision tree generalizes the data. Then, this decision tree can be mapped to a set of production rules by writing one rule for each path of the tree, in particular.

- If a patient has swollen glands, the diagnosis is strep throat.
- If a patient does not have swollen glands and does not have a fever, the diagnosis is an allergy.
- If a patient does not have swollen glands and does have a fever, the diagnosis is a cold.

The decision tree describes how to accurately diagnose a patient by only considering whether the patient has swollen glands and a fever. The attributes sore throat, congestion, and headache do not play a role in determining this diagnosis.

ID3 is the most widely cited decision tree construction algorithm and several variations and enhancements have been proposed [24-26]. There are also a related branch of approaches that are not based on ID3 [27-29].

3. Methodology

The procedures used for the implementation of the proposed system are shown in Fig. 2. There are five principal modules in the system, including data collection, data transformation, data clustering, decision trees construction and decision trees simplification.

Data collection module: This module is used to collect data for analysis in the study. It is effectively a database.

Data transformation module: It is used to preprocess the raw data for underlying analysis.

Data clustering module: The module is to cluster data into smaller groups according to fuzzy ART2 model. This module helps reduce the complexity of constructing the decision trees.
Decision trees construction module: Data in the same cluster are used to create the decision trees that can be further transformed into semantically fuzzy rules.

Decision trees simplification module: The purpose is to further prune the trees.

A. Preprocessing

The ART2 input patterns are fuzzified and the clustered results are used to find the association rules among data. Assume that each input pattern contains \( n \) attributes \((X_1, X_2, \ldots, X_n)\) and a membership function \( \mu_A \) of set \( A \) maps each attribute value to a real number in the interval \([0,1]\). For convenience, the fuzzy membership function is denoted \( \mu_s \) where the subscript \( f \) indicates the corresponding fuzzy subset. The set of fuzzy membership functions for all attributes is denoted by \( S \). The membership functions labels for each attribute are normally determined by the distribution of the attribute values.

The ART2 input pattern are represented by new patterns \((\mu_{i1}, \mu_{i2}, \ldots, \mu_{in}, \mu_{a1}, \mu_{a2}, \ldots, \mu_{an})\) where \( n \) is the number of attributes in the dataset, and \((\mu_{i1}, \mu_{i2}, \ldots, \mu_{in})\) represent the membership functions of the \( i \)th attribute in fuzzy subset \( S \). To better facilitate the discovery of association rules, linguistic terms are used instead of membership functions.

B. Fuzzy ART2 Neural Network

The following paragraphs describe the proposed fuzzy neural network.

1) Network Structure: After the fuzzy inputs have been extracted, the fuzzy ART2 clusters the patterns. The fuzzy ART2 can be described as follows:

- **Input layer**: The input layer consists of short term memories (STM).
- **Weight layer**: The weight layer consists of bottom-up \( b_l \) and top-down \( t_l \) weights, called long term memories (LTM).
- **Output layer**: The output layer is used to express the clustering results for the given data.

2) Learning Procedures:

The fuzzy ART2 learning procedure consists of the following four steps:

- **Step 1**: The fuzzy vector is input into the input layer.
- **Step 2**: The distance between the bottom-up weights and fuzzy inputs are calculated and the shortest distance is identified.

- **Step 3**: If the shortest distance fails a vigilance test (a threshold defined to test the similarity between a new pattern and a cluster centroid), a new node is created with its weights equal to the fuzzy inputs. If a cluster wins the vigilance test, the centroid of the cluster is adjusted to adopt the new input.

- **Step 4**: The process is repeated until all the data are clustered.

If all winners fail the vigilance test a new cluster is created and the corresponding weights are added. If the state is “resonance”, the current fuzzy input is assigned to this cluster by modifying the corresponding weights.

The fuzzy ART2 is performed before discovering the decision tree rules to divide the volume of data from a potentially large database into smaller and more manageable subsets. Moreover, since patterns in a cluster have similar characteristics, the time taken to construct the decision tree for the subset will be shorter than if the process is applied to the full dataset.

C. Finding Decision Tree Rules

After all patterns of the original data are grouped into clusters, the data classification algorithms are used to find the decision tree rules for each cluster.

The fuzzy decision tree [32] is an extension of the traditional decision tree and an effective method for extracting knowledge given uncertain classification problems. Fuzzy set theory is applied to the dataset. Tree growing and pruning are combined to determine the tree structure.

ID3 selects the test attribute based on the information gain which is computed according to the probability of ordinary data. The algorithm proposed herein is similar to ID3, but its gain is computed according to the membership degrees.

Assume that we have a dataset \( D \), where each datum has \( l \) numerical attributes \( A_1, A_2, \ldots, A_l \), one class set \( C = \{C_1, C_2, \ldots, C_n\} \) and fuzzy subset \( D_{C_l} \) for the attribute \( A_l \) (assuming attribute \( A_l \) has \( n \) different values). Let \( D_{C_l} \) be a fuzzy subset in \( D \) whose class is \( C_l \) and \( |D| \) the sum of the membership values in the dataset \( D \). Then, the fuzzy decision tree is generated as follows:

(a) Generate the root node \( T \).
(b) If a node \( t \) with a fuzzy dataset \( D \) satisfies the following conditions, then it is a leaf node and assigned the majority class in \( D \).

\[
\frac{|D_{C_k}|}{|D|} \geq \theta, \quad (1)
\]
(2). The sum of the membership values in a fuzzy dataset is less than a threshold $\theta_n$, that is, 
$$ |D_{C_i}| \leq \theta_n. \quad (2) $$
(3). There are no attributes for further classification.
(c) If it does not satisfy the above conditions, it is not a leaf and the test node is generated as follows:
(1). For $A_i$'s ($i = 1, 2, \ldots, n$), calculate the information gains $G(A_i, D)$, described below, and select the maximizing test attributes $A_{\text{max}}$.
(2). Divide $D$ into fuzzy subsets $D_1, D_2, \ldots, D_m$ according to $A_{\text{max}}$ and generate new nodes $t_1, t_2, \ldots, t_m$ for the fuzzy subsets $D_1, D_2, \ldots, D_m$.
(3). Replace $D$ by $D_j$ ($j = 1, 2, \ldots, m$) and repeat recursively from (b).

The information gain $G(A, D)$ for attribute $A$ with a fuzzy dataset $D$ is defined by
$$ G(A, D) = I(D) - E(A, D), $$
where
$$ I(D) = -\sum_{k=1}^{n} p_k \log_2 p_k, \quad E(A, D) = \sum_{j=1}^{m} p_j I(D_{A_j}). $$

\[ p_j = \frac{|D_{A_j}|}{|D|}, \quad p_k = \frac{\sum_{h=1}^{\infty} |D_{C_h}|}{|D|} \]

where $|D_{C_h}|$ is the sum of membership values in a fuzzy subset $D_{C_j}$ for the attribute $A_j$.

Eq.(1) indicates that if all data belong to a single cluster, then there is no need for further split on the decision node. Eq.(2) is used to check whether a cluster $C_k$ contains only few data. In our experiments, we set $\theta_1 = 1$ and $\theta_0 = 0$.

The following sleeping quality measurement example illustrates one cycle of the algorithm with the data in Table 1. Given a fuzzy dataset $D$ where MV is the membership value of each data set. There are two distinct classes with values G (Good) and B (Bad). A (root) node T is created for the samples in $D$. To find the splitting criterion for these samples, the information gains of the four test attributes are computed according to the fuzzy values and the results compared.

First, the expected information needed to classify a sample in $D$ from Eq.(3) is
$$ I(D) = -\frac{9}{20} \log_2 \frac{9}{20} - \frac{11}{20} \log_2 \frac{11}{20} = 0.99 \text{ bits}. $$
Next, the expected information requirement for each attribute is found. For each linguistic term in each attribute, the total fuzzy values of each linguistic term in the same class G or B are found by applying Eq.(4). For attribute “clothes” in Table 2 shows the total fuzzy values of the linguistic terms in each class. Note that the first figure in the Table 2, “much” (2-5), means that two data belong to the class G while 5 data in class B for clothes with an attribute value of “much.” The expected information needed to classify a sample in $D$ if the examples are partitioned according to “clothes” is:
$$ E(\text{Clothes}, D) = \frac{5}{13.7} I(D_{\text{clothes}}) + \frac{4.3}{13.7} I(D_{\text{clothes}}) + \frac{4.4}{13.7} I(D_{\text{clothes}}) = 0.54 \text{ bits}. $$
Therefore, the information gained by branching on attribute “clothes” is:
$$ G(\text{clothes}, D) = I(D) - E(\text{clothes}, D) = 0.45 \text{ bits}. $$

<table>
<thead>
<tr>
<th>MV</th>
<th>Temp</th>
<th>MV</th>
<th>Humidity</th>
<th>MV</th>
<th>Noise</th>
<th>MV</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>much</td>
<td>1.0</td>
<td>high</td>
<td>0.6</td>
<td>prodigious</td>
<td>0.2</td>
<td>no</td>
<td>0.9</td>
</tr>
<tr>
<td>much</td>
<td>0.5</td>
<td>high</td>
<td>0.9</td>
<td>prodigious</td>
<td>0.7</td>
<td>noisy</td>
<td>0.3</td>
</tr>
<tr>
<td>much</td>
<td>0.7</td>
<td>high</td>
<td>0.3</td>
<td>prodigious</td>
<td>0.5</td>
<td>medium</td>
<td>0.6</td>
</tr>
<tr>
<td>normal</td>
<td>0.4</td>
<td>high</td>
<td>0.5</td>
<td>prodigious</td>
<td>0.8</td>
<td>no</td>
<td>0.7</td>
</tr>
<tr>
<td>normal</td>
<td>0.8</td>
<td>high</td>
<td>0.2</td>
<td>prodigious</td>
<td>1.0</td>
<td>medium</td>
<td>0.2</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.6</td>
<td>moderate</td>
<td>0.7</td>
<td>prodigious</td>
<td>0.9</td>
<td>no</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.7</td>
<td>comfortable</td>
<td>1.0</td>
<td>prodigious</td>
<td>0.6</td>
<td>medium</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.7</td>
<td>high</td>
<td>comfortable</td>
<td>0.8</td>
<td>no</td>
<td>0.3</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>1.0</td>
<td>high</td>
<td>comfortable</td>
<td>0.5</td>
<td>noisy</td>
<td>0.2</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.9</td>
<td>moderate</td>
<td>0.6</td>
<td>prodigious</td>
<td>0.9</td>
<td>no</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.2</td>
<td>moderate</td>
<td>0.7</td>
<td>prodigious</td>
<td>0.4</td>
<td>medium</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.6</td>
<td>comfortable</td>
<td>0.4</td>
<td>comfortable</td>
<td>0.2</td>
<td>no</td>
</tr>
<tr>
<td>enough</td>
<td>high</td>
<td>0.3</td>
<td>moderate</td>
<td>0.9</td>
<td>comfortable</td>
<td>0.7</td>
<td>medium</td>
</tr>
<tr>
<td>much</td>
<td>high</td>
<td>0.7</td>
<td>moderate</td>
<td>0.3</td>
<td>comfortable</td>
<td>0.6</td>
<td>medium</td>
</tr>
<tr>
<td>much</td>
<td>high</td>
<td>1.0</td>
<td>moderate</td>
<td>1.0</td>
<td>comfortable</td>
<td>1.0</td>
<td>noisy</td>
</tr>
<tr>
<td>normal</td>
<td>high</td>
<td>1.0</td>
<td>moderate</td>
<td>0.8</td>
<td>prodigious</td>
<td>0.2</td>
<td>noisy</td>
</tr>
<tr>
<td>normal</td>
<td>high</td>
<td>0.8</td>
<td>moderate</td>
<td>0.9</td>
<td>prodigious</td>
<td>1.0</td>
<td>medium</td>
</tr>
<tr>
<td>normal</td>
<td>high</td>
<td>0.3</td>
<td>moderate</td>
<td>0.5</td>
<td>comfortable</td>
<td>0.9</td>
<td>no</td>
</tr>
<tr>
<td>normal</td>
<td>high</td>
<td>0.9</td>
<td>moderate</td>
<td>0.3</td>
<td>prodigious</td>
<td>0.3</td>
<td>noisy</td>
</tr>
<tr>
<td>normal</td>
<td>high</td>
<td>1.0</td>
<td>high</td>
<td>comfortable</td>
<td>0.8</td>
<td>medium</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Similarly, the information gained by branching on the three remaining attributes can be calculated in the same
way with the attribute “clothes”. Finally, the results of
the gain in information by branching on all four attrib-
utes are displayed in decreasing bit values in Table 3.

The attribute “clothes” in Table 3 has the highest in-
formation gain and becomes the splitting attribute at
the root node T of the fuzzy decision tree. The root node
“clothes” has three linguistic terms “much”, “normal”
and “enough”, thus three leaves are grown from this
node. Proceeding the same way for other attributes the
final fuzzy decision tree in Fig. 3 was obtained.

Table 3. Information gained by branching on four attributes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>clothes</td>
<td>0.45</td>
</tr>
<tr>
<td>humidity</td>
<td>0.08</td>
</tr>
<tr>
<td>noise</td>
<td>0.06</td>
</tr>
<tr>
<td>temperature</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Figure 3. Final decision tree using fuzzy ID3 algorithm.

4. Experimental Results and Analyses

Benchmark data were used to illustrate the effective-
ness of the proposed model. The data sets are available
from the UCI machine learning repository [33].

The first experiment involves Haberman’s Survival
problem. The Harbeman’s Survival dataset consists of
306 samples. Each sample comprises four attributes ab-
Abbreviated Attr1, Attr2, Attr3, and Attr4. Table 4 lists
the linguistic terms used for these four attributes in each
pattern for this experiment. The survival status attribute
value is clearly indicated in the medical data, not
self-defined information. Besides, either one or two in-
dicates whether or not the patient has survived five or
more years.

The second experiment addressed the Blood Transfu-
sion problem. There are 748 samples in this dataset
where each sample contains five attributes. The classi-
fying attribute is a binary variable representing whether
the donor should donate blood at a certain time. The
linguistic terms for the attributes are listed in Table 5.

Table 4. Abbreviations of the patient parameters.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>No.</th>
<th>Parameter</th>
<th>Linguistic Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age of patient</td>
<td>3</td>
<td>Ma</td>
<td>Middle_Aged</td>
</tr>
<tr>
<td>4</td>
<td>Se</td>
<td>Senior</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Ls</td>
<td>Little_Senior</td>
<td></td>
</tr>
<tr>
<td>Patient’s year</td>
<td>1</td>
<td>Ol</td>
<td>Old</td>
</tr>
<tr>
<td>of operation</td>
<td>2</td>
<td>Av</td>
<td>Average</td>
</tr>
<tr>
<td>3</td>
<td>Re</td>
<td>Recent</td>
<td></td>
</tr>
<tr>
<td>Number of</td>
<td>1</td>
<td>Vs</td>
<td>Very_Small</td>
</tr>
</tbody>
</table>
| positive auxil-
| 2               | Sm  | Small       |                  |
| iary nodes      | 3               | Me          | Medium           |
| detected        | 4               | Bi          | Big              |
| 5               | Vb  | Very_Big    |                  |
| Survival status | 1   | Hi          | High             |
| 2               | Lo  | Low         |                  |

Table 5. Abbreviations of the donor parameters.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>No.</th>
<th>Parameter</th>
<th>Linguistic Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recency - months</td>
<td>1</td>
<td>Vs</td>
<td>Very_Small</td>
</tr>
<tr>
<td>since last</td>
<td>2</td>
<td>Sm</td>
<td>Small</td>
</tr>
<tr>
<td>donation</td>
<td>3</td>
<td>Me</td>
<td>Medium</td>
</tr>
<tr>
<td>4</td>
<td>Bi</td>
<td>Big</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Vb</td>
<td>Very_Big</td>
<td></td>
</tr>
<tr>
<td>Frequency - total</td>
<td>1</td>
<td>Te</td>
<td>Few</td>
</tr>
<tr>
<td>number of</td>
<td>2</td>
<td>Me</td>
<td>Medium</td>
</tr>
<tr>
<td>donation</td>
<td>3</td>
<td>Ma</td>
<td>Many</td>
</tr>
<tr>
<td>Monetary - total</td>
<td>1</td>
<td>Ti</td>
<td>Too_Little</td>
</tr>
<tr>
<td>blood donated in</td>
<td>2</td>
<td>Li</td>
<td>Little</td>
</tr>
<tr>
<td>c.c.</td>
<td>3</td>
<td>Av</td>
<td>Average</td>
</tr>
<tr>
<td>4</td>
<td>Mu</td>
<td>Much</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Tm</td>
<td>Too_Much</td>
<td></td>
</tr>
<tr>
<td>Time - months</td>
<td>1</td>
<td>Vr</td>
<td>Very_Recent</td>
</tr>
<tr>
<td>since first</td>
<td>2</td>
<td>Re</td>
<td>Recent</td>
</tr>
<tr>
<td>donation</td>
<td>3</td>
<td>Mo</td>
<td>Moderate</td>
</tr>
<tr>
<td>4</td>
<td>Lg</td>
<td>Long</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Vl</td>
<td>Very_Long</td>
<td></td>
</tr>
<tr>
<td>A binary variable</td>
<td>1</td>
<td>Hi</td>
<td>High</td>
</tr>
<tr>
<td>representing</td>
<td>2</td>
<td>Lo</td>
<td>Low</td>
</tr>
<tr>
<td>whether he/she</td>
<td>307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>donated blood</td>
<td>2007</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A. Haberman’s Survival Problem

The vigilance value in the ART2 model dominates the
clustering results that in turn affect the number of pat-
terns assigned to each cluster. Depending on the data-
base size and the properties of the data, it is hard to de-
cide a suitable number of clusters before clustering. In
the first experiment, some mid-size clusters, such as 3 to
7, are tested for the medical data. Small clusters are
checked to see whether they are outliers or not. For ex-
ample, we found that one pattern (83, 58, 2, 2) is nu-
merically distant from the remaining data. Therefore,
only 305 patterns are used in the simulations to avoid the
overhead of processing outliers.

Experiment 1 started by fuzzifying the input patterns
using the membership function labels 3, 3, 3 and 2 for
the four attributes (case HS-1 with the vigilance value
is set to 3.6). The membership functions of the four attrib-
utes are also plotted in Fig. 4.
The overlapping range and shape of the membership functions may affect the decision rules. Although these parameters can be further optimized, this is not the focus of this study. Here, all membership functions are determined empirically.

Another objective of this experiment was to examine the effect of the number of membership functions on building decision trees. The number of membership functions for the patient age attribute and the number of positive auxiliary nodes attribute are changed to 5 membership functions as shown in Fig. 7 (case HS-2, the vigilance value is set to 3.4).

- The entire tree contains the leaf MeMaAvLo (indicated with an arrow in Fig. 6). It means the patient did not survive five or more years. The classification result (Lo) is contradictory to the evidence given in pattern 53.

In this case, the dataset is clustered into 5 clusters (larger than that of case HS-1) where the numbers of patterns are 78, 66, 57, 85 and 19. In two cases the full advantage of fuzzy ART2 is taken in finding decision rules.

Figure 4. Linguistic terms for the four attributes for case HS-1.

Figure 5. The decision trees for case HS-1: (a) Cluster 1, (b) Cluster 2 and (c) Cluster 3.

Figure 6. The decision tree for the original dataset in case HS-1.

Figure 7. Linguistic terms of attribute 1 and attribute 3 for case HS-2.
tree rules. The dataset is split into smaller parts to reduce the computational effort. Table 6 lists the results of the first experiment including the data ratio and reduction percentage of erroneous classifications in the clusters for the two cases.

To compare the whole tree with the other smaller trees of clusters, we use all patterns in each cluster to check for erroneously clustered patterns. Table 6 also shows that the full tree always contains more erroneously clustered patterns than the small trees. For example, cluster 1 has five erroneously clustered patterns in the full tree while the small tree only has one in case HS-1. The error classification rates are reduced by 47.1% and 58.1% in cases HS-1 and HS-2, respectively.

The trees in case HS-2 are more complicated than case HS-1 because increasing the number of linguistic terms for attributes also incurs an increase in branches for each attribute. The results in Table 6 show that there are more erroneously clustered patterns in the full tree than in the small trees. Cluster 5 has no erroneously clustered pattern (the best case).

The decision tree and the corresponding classification rules are further simplified [34]. The trimmed trees from Fig. 5 are shown in Fig. 8. It is easier to search for decision rules in these reduced trees. For example, Fig. 8(a) has only three leaves in the reduced tree and the rules are formed: SeHi, MaHi and YoHi. Thus, if any pattern belongs to cluster 1, it will be classified to attribute Hi.

Table 6. Results of experiment 1.

<table>
<thead>
<tr>
<th>Cluster No. of patterns</th>
<th>Data ratio (%)</th>
<th>Full tree erroneous</th>
<th>From cluster erroneous</th>
<th>Reduction percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-1</td>
<td>131</td>
<td>43.0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>115</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>59</td>
<td>19.3</td>
<td>10</td>
</tr>
<tr>
<td>Case 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS-2</td>
<td>78</td>
<td>25.6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>66</td>
<td>21.6</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>57</td>
<td>18.7</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>85</td>
<td>27.9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>19</td>
<td>6.2</td>
<td>13</td>
</tr>
</tbody>
</table>

B. Blood Transfusion Service Center Database

The proposed method was also applied to the Blood Transfusion Service Center database. First, seven outliers were discarded and 741 patterns were used. Two cases were considered with the vigilance values set to 3.9. The inputs of the first case, BT-1, were fuzzified using the number of membership functions of 3, 3, 3, 3 and 2 for the five attributes. The simplified trees from clusters 1, 2, 3, 4 and full tree in case BT-1 are shown in Fig. 9. For the second case, BT-2, the recency (months since last donation attribute), the monetary (total blood donated in c.c.) attribute and time (months since first donation) were respectively fuzzified using five membership functions.

Table 7 presents the grouping with distributions of patterns achieved with the fuzzy ART2 for the two cases. The datasets of the two cases are grouped into 4 and 5 clusters, respectively. The decision tree rules help make decisions about whether someone should donate blood during a certain period.

To demonstrate the efficiency of the fuzzy ART2 algorithm, the erroneously clustered patterns are also extracted and the results are listed in Table 7. The results show that when processing the whole tree more errone-
ously clustered patterns appear than for the small trees. The error classification rates are reduced by 34.5% and 92.9% in cases BT-1 and BT-2, respectively. Finally, the resulting trees are simplified as described previously.

Although there are 4 and 5 attributes in the Haberman’s Survival database and Blood Transfusion Service Center database, respectively, the proposed method will not be deterred its applications to multi-attribute databases. The proposed method can be easily applied to construct the decision trees from multi-attribute databases.

Finally, the resulting trees are simplified as described previously. Although there are 4 and 5 attributes in the Haberman’s Survival database and Blood Transfusion Service Center database, respectively, the proposed method will not be deterred its applications to multi-attribute databases. The proposed method can be easily applied to construct the decision trees from multi-attribute databases.

Table 7. Results for experiment 2.

<table>
<thead>
<tr>
<th>Cluster No. of patterns Data ratio (%)</th>
<th>Full tree From cluster Reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT-1 Case 1 104 14.0 28 19 32.1</td>
<td></td>
</tr>
<tr>
<td>2 252 34.0 0 0 0 0</td>
<td></td>
</tr>
<tr>
<td>3 229 30.8 1 0 100</td>
<td></td>
</tr>
<tr>
<td>4 157 21.2 0 0 0</td>
<td></td>
</tr>
<tr>
<td>BT-2 Case 1 91 12.3 22 0 100</td>
<td></td>
</tr>
<tr>
<td>2 259 35.0 28 0 100</td>
<td></td>
</tr>
<tr>
<td>3 297 40.1 1 0 100</td>
<td></td>
</tr>
<tr>
<td>4 44 5.9 5 4 20</td>
<td></td>
</tr>
<tr>
<td>5 50 6.7 0 0 0</td>
<td></td>
</tr>
</tbody>
</table>

5. Conclusions and Future Work

A novel approach is proposed for finding decision tree rules by combining fuzzy ART2 and fuzzy ID3. First, a fuzzy model and ART2 neural network is used for clustering the fuzzified dataset into groups with similar properties. Next, decision tree rules are mined using the fuzzy ID3.

Two medical datasets have been explored to validate the effectiveness of the proposed approach, namely the Haberman’s Survival database for grouping patients who have undergone surgery for breast cancer into groups with similar properties, and the Blood Transfusion Service Center database for clustering blood donors into groups. The experiments confirm that the erroneous classification rates are significantly reduced by processing simplified decision trees. Next, the experimental results show that the discovered decision tree rules from individual cluster are more efficient than those based on all of the data. The number of erroneously clustered patterns in the decision tree of individual cluster is smaller than for the entire tree. Applications of the proposed method are not limited to medical databases. As long as the accuracy and efficiency of data classification are concerned, the presented work is applicable to the multi-attribute databases.

Although the ID3 algorithm is very popular in the field of machine learning it is associated with several drawbacks. For example, it tends to select attributes that have more attribute values as the expanding attribute in the process of building decision tree, but the selected attribute is often not the one that contributes most to the classification. Future work should thus focus on improving the ID3 algorithm and the classification accuracy.

Acknowledgments

This work was supported in part by National Science Council, Taiwan under Grant NSC100-2221-E-027-110 and in part by Joint project between National Taipei University of Technology and Mackay Memorial Hospital under Grants NTUT-MMH-99-03 and NTUT-MMH-100-09.

References

[35] R. Jin and G. Agrawal, “Efficient decision tree construction on streaming data,” in Proc. of the Ninth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Washington, D.C.
Yo-Ping Huang et al.: Improving Classifications of Medical Data Based on Fuzzy ART2 Decision Trees


Yo-Ping Huang received his Ph.D. in electrical engineering from Texas Tech University, Lubbock, TX, USA. He is currently a Professor in the Department of Electrical Engineering at National Taipei University of Technology (NTUT), Taiwan. He also serves as CEO of Joint Commission of Technological and Vocational College Admission Committee in Taiwan. He was the secretary general at NTUT, Chairman of IEEE CIS Taipei Chapter, and Vice Chairman of IEEE SMC Taipei Chapter. He was professor and Dean of College of Electrical Engineering and Computer Science, Tatung University, Taipei, before joining NTUT. His research interests include medical knowledge mining, intelligent control systems, handheld device application systems design. Prof. Huang is a senior member of the IEEE and a fellow of the IET.

Shin-Liang Lai received BS and MS degrees in mathematics from the National Taipei University of Education, Taipei, Taiwan, in 1998 and 2002, respectively. He is currently a researcher and Ph.D. candidate at the department of computer science and engineering, Tatung University, Taipei, Taiwan. His research interests include content-based music information retrieval, data mining, and fuzzy inference system.

Frode Eika Sandnes received a B.Sc. in computer science from the University of Newcastle upon Tyne, U.K., and a Ph.D. in computer science from the University of Reading, U.K. Sandnes is a Professor in the Institute of Information Technology, Faculty of Technology, Art and Design at Oslo and Akershus University College of Applied Sciences, Norway. He currently serves as pro-Rector of research and internationalisation at Oslo and Akershus University College of Applied Sciences. His research interests include error-correction, intelligent systems, human computer interaction and universal design.

Shen-Ing Liu received her M.D. degree from the China Medical University, Taiwan, and Ph.D. degree in psychiatric epidemiology from the Institute of Psychiatry, King’s College London, University of London, U.K. She is currently an Associate Professor in the Department of Medicine at Mackay Medi-cal College, Taiwan and also serves as an attending psychiatrist in Mackay Memorial Hospital, Taipei, Taiwan. Her research interests include epidemiology, clinical psychiatry and interventional study.