FUZZY DECISION TREE BASED ON FUZZY ROUGH SETS

G. JAYA SUMA and GANKIDI SANTHOSH REDDY

1G. Jaya Suma
Associate Professor
Dept of Computer Science and Engineering, GIT,
GITAM University, Visakhapatnam, India.
yasuma_g@yahoo.co.in
98491-98738

2Gankidi Santhosh Reddy
M.Tech
Dept of Computer Science and Engineering, GIT,
GITAM University, Visakhapatnam, India.
santhoshreddy.g@gmail.com
88013-03495

Abstract

Decision trees are very successful in building classifier with nominal attributes. They are less effective for ordinal and continuous valued attributes due to sharp split points. These sharp split points may lead to misclassification, which can be handled by Fuzzy decision trees in a better way. Many existing algorithms are used for the development of fuzzy decision tree. In all these algorithms fuzzification should be done by partitioning the data into fuzzy sets. For generation of fuzzy decision tree, attributes need to be fuzzified. In this context, selection of fuzzy attributes is important in construction of fuzzy decision trees. Different fuzzy attributes have different influences on decision making. Some of them are more important than others. In this regard, fuzzy classification entropy and classification ambiguity both are essentially use the ratio of uncertainty to measure the significance of fuzzy attributes. In this paper, we use fuzzy rough technique, in which the attributes are selected by using the significance of fuzzy attributes with respect to fuzzy class label. The fuzzy rough approach is experimented with several UCI datasets like iris and glass datasets for testing performance.

Keywords: Fuzzification, Fuzzy attributes, Fuzzy class label, Fuzzy decision trees, Fuzzy rough technique.

1 Introduction

Classification is an important concept in data mining. It has been studied extensively by the machine learning community as a possible solution to the knowledge acquisition or knowledge extraction problem. The input to the classifier construction algorithm is a training set of records, each of which is tagged with a class label. A set of attribute values defines each record. Attributes with discrete domains are referred to as categorical, while those with ordered domains are referred to as numeric. The goal is to induce a model or description for each class in terms of the attributes. The model is then used by the classifier to classify future records whose classes are unknown.

Classification has been successfully applied to several areas like medical diagnosis, weather prediction, credit approval, customer segmentation and fraud detection. Among the techniques developed for classification, popular ones include bayesian classification (Cheeseman et al., 1988), neural networks (Bishop, 1995; Ripley, 1996), genetic algorithms (Goldberg, 1989) and decision trees (Breiman et al., 1984): The popularity of decision trees is due to conceptual transparency, inexpensive computation, ease of interpretation and its ability to represent knowledge in form of rules unlike neural networks. There are several reasons for this (Breiman et al.,
neural networks training can take large amounts of time and thousands of iterations, inducing decision trees is efficient and is thus suitable for large training sets. Also, decision tree generation algorithms do not require additional information besides that already contained in the training data (e.g., domain knowledge or prior knowledge of distributions on the data or classes). Finally, decision trees display good classification accuracy compared to other techniques.

Classification decision tree suffers with sharp split points. To improve classification accuracy of continuous attribute fuzzy representation is introduced to the classical decision trees called fuzzy decision trees.

In ID3 algorithm (Ross Quinlan 1983) entropy used to determine which node to split. ID3 has highly unstable classifiers with respect to minor perturbation in training data and data may be over-fitted or over-classified, if a small sample is tested. Only one attribute at a time is tested for making a decision. Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum in ID3. Fuzzy logic brings in an improvement of these aspects due to the elasticity of fuzzy sets formalism. In order to overcome it, Fuzzy ID3 proposed by Shaw and Yuan (1995) uses classification ambiguity as a measure to select the expanded node in the fuzzy decision tree. However entropy and ambiguity are two types of uncertainties.

Rough sets developed by Pawlak (1982) are useful tools to deal with uncertainties. Fuzzy rough sets are combines the fuzzy sets and rough sets, deals with both fuzziness and uncertainties.

Given a Fuzzy System (FS), there are many fuzzy attributes and each attributes have different significance to classification. In this paper, we use fuzzy rough method, which uses significance instead of fuzzy entropy or classification ambiguity to select the most important attribute recursively to generate fuzzy decision tree.

2 Fuzzy System and Fuzzy Rough Sets

2.1 Fuzzy System

Fuzzy system is a 4-tuple FS={ U, C, D, V, f} where U={O1, O2, O3,...,On} is a set of finite objects and each Oi={C1, C2, C3, ..., Cn}; C={C1, C2, C3, ..., Cn} where Ci (1≤i≤n) represents a fuzzy attribute which consists of a set of fuzzy linguistic terms FLTi={C1, C2, C3, ..., Cn} (1≤i≤n); D denotes fuzzy class label attribute with a set of fuzzy linguistic terms FLT D={D1, D2, ..., Dn}. Each fuzzy linguistic term Cj (1≤i≤n; 1≤j≤k) or Ci (1≤i≤m) is considered as a fuzzy set on U, represented as

\[
C_{ij} = \frac{c_{ij}^{(1)}}{\alpha_1} + \frac{c_{ij}^{(2)}}{\alpha_2} + \ldots + \frac{c_{ij}^{(N)}}{\alpha_N}
\]

and

\[
D_{ij} = \frac{d_{ij}^{(1)}}{\alpha_1} + \frac{d_{ij}^{(2)}}{\alpha_2} + \ldots + \frac{d_{ij}^{(N)}}{\alpha_N}
\]

where c_{ij}^{(p)} (1≤i≤n; 1≤j≤k; 1≤p≤N) is the conditional membership degree, while d_{ij}^{(p)} (1≤p≤N) is the class membership degree; \( V = \bigcup_{j=1}^{n} V_{a_j} \) where \( V_{a_j} \) is the domain value of the attribute \( a_j \); f: \( U \times C \rightarrow V \) is a function \( f(o_i, c_j) = c_{ij} \) (1≤i≤n; 1≤j≤n).

2.2 Fuzzy Rough Sets

Fuzzy rough sets developed originally by D. Dubois integrate together the concepts of vagueness and indiscernibility. Here, we review some preliminaries related to fuzzy rough method.

**Definition 2.2.1** Let U be a given universe. A partition p={F1, F2, ..., Fn} of U is a fuzzy partition if and only if the following to requirements hold,

1. \( \forall x_i \in U, \forall f_j \in p, \mu_{f_j}(x_i) \leq 1; \)
2. \( \forall x_i \in U, \exists f_j \in p, \mu_{f_j}(x_i) \geq 0; \)

where \( \mu_{f_j}(x_i) \) denotes the membership degree to which \( x_i \) belongs \( F_j \).

**Definition 2.2.2** Let U be a given universe. R is a fuzzy equivalence relation over U if the following four requirements hold:

1. R is a fuzzy relation on U;
2. R is reflexive i.e. \( R(x, x) = 1, \forall x \in U; \)
(3) $R$ is symmetric, i.e. $R(x, y) = R(y, x)$, $\forall x, y \in U$;
(4) $R$ is transitive, i.e. $R(x, z) \geq \min \{R(x, y), R(y, z)\}$, $\forall x, y, z \in U$;

**Table 1 Fuzzy System**

<table>
<thead>
<tr>
<th>No.</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>...</th>
<th>$C_n$</th>
<th>$D_1$</th>
<th>$D_2$</th>
<th>...</th>
<th>$D_m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_1$</td>
<td>$c_1^{(1)}$</td>
<td>$c_2^{(1)}$</td>
<td>...</td>
<td>$c_{k_1}^{(1)}$</td>
<td>$c_1^{(1)}$</td>
<td>$c_2^{(1)}$</td>
<td>...</td>
<td>$c_{k_2}^{(1)}$</td>
</tr>
<tr>
<td>$O_2$</td>
<td>$c_1^{(2)}$</td>
<td>$c_2^{(2)}$</td>
<td>...</td>
<td>$c_{k_1}^{(2)}$</td>
<td>$c_1^{(2)}$</td>
<td>$c_2^{(2)}$</td>
<td>...</td>
<td>$c_{k_2}^{(2)}$</td>
</tr>
<tr>
<td>$O_n$</td>
<td>$c_1^{(N)}$</td>
<td>$c_2^{(N)}$</td>
<td>...</td>
<td>$c_{k_1}^{(N)}$</td>
<td>$c_1^{(N)}$</td>
<td>$c_2^{(N)}$</td>
<td>...</td>
<td>$c_{k_2}^{(N)}$</td>
</tr>
</tbody>
</table>

**Definition 2.2.3** Let $U$ be a given universe. $R$ is a fuzzy equivalence relation over $U$. The fuzzy equivalence class $[x]_R$ is defined by

\[
\mu_{[x]_R}(y) = \mu_R(x, y)
\]

**Definition 2.2.4** Let $U$ be a given universe. $X$ and $P$ are two fuzzy sets on $U$. The fuzzy $P$-lower and $P$-upper approximations are defined as follows

\[
\mu_{\text{PL}}(F_i) = \inf_{x \in U} \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad (i = 1, 2, ..., n)
\]

\[
\mu_{\text{PU}}(F_i) = \sup_{x \in U} \min\{\mu_{F_i}(x), \mu_X(x)\} \quad (i = 1, 2, ..., n)
\]

Where $F_i \in U/P \ (1 \leq i \leq m)$ denotes a fuzzy equivalence class and $U/P$ denotes the fuzzy partition of $U$ with respect to $P$.

**Definition 2.2.5** Let $U$ be a given Universe, $X$ and $P$ be two fuzzy sets on $U$, $U/P$ be a fuzzy partition of $U$, for a given $x \in U$, the fuzzy $P$-lower approximation and the fuzzy $P$-upper approximation of $X$ are defined as follows

\[
\mu_{\text{PL}}(x) = \sup_{F \in U/P} \min\{\mu_x(x), \inf_{y \in U} \max\{1 - \mu_{F}(y), \mu_X(y)\}\}
\]

\[
\mu_{\text{PU}}(x) = \sup_{F \in U/P} \min\{\mu_x(x), \sup_{y \in U} \min\{\mu_{F}(y), \mu_X(y)\}\}
\]
The tuple \((PX, \overline{PX})\) is called a fuzzy rough set.

### Table 2 A sample fuzzy system with 9 instances

<table>
<thead>
<tr>
<th>Number</th>
<th>Sepal length</th>
<th>Sepal Width</th>
<th>Petal length</th>
<th>Petal width</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>short</td>
<td>long</td>
<td>short</td>
<td>Long</td>
<td>setosa</td>
</tr>
<tr>
<td>2</td>
<td>short</td>
<td>long</td>
<td>0.714</td>
<td>0.285</td>
<td>versicolor</td>
</tr>
<tr>
<td>3</td>
<td>short</td>
<td>0.142</td>
<td>0.857</td>
<td>1</td>
<td>0.9956</td>
</tr>
<tr>
<td>4</td>
<td>short</td>
<td>0.142</td>
<td>0.85</td>
<td>0.1266</td>
<td>0.00341</td>
</tr>
<tr>
<td>5</td>
<td>short</td>
<td>0.214</td>
<td>0.571</td>
<td>0.074</td>
<td>0.994</td>
</tr>
<tr>
<td>6</td>
<td>short</td>
<td>0.272</td>
<td>0.727</td>
<td>1</td>
<td>0.9109</td>
</tr>
<tr>
<td>7</td>
<td>short</td>
<td>0.581</td>
<td>0.4179</td>
<td>0.0218</td>
<td>0.978</td>
</tr>
<tr>
<td>8</td>
<td>short</td>
<td>0.714</td>
<td>0.285</td>
<td>1</td>
<td>0.904</td>
</tr>
<tr>
<td>9</td>
<td>short</td>
<td>0.142</td>
<td>0.85</td>
<td>0.1266</td>
<td>0.00341</td>
</tr>
</tbody>
</table>

3. **Fuzzy Decision Tree Construction using Fuzzy Rough Concept**

3.1 **Description of Fuzzy Rough Concept**

In this section, we use significance to select expanded attribute in the generation of fuzzy decision tree. As each attribute has different importance, significance of the fuzzy attribute with respect to the fuzzy class label is used. The significance of fuzzy attribute \(C\) with respect to fuzzy class label \(D\) is obtained as follows:

\[
\tau_C(D) = \frac{\sum_{x \in U} \mu_{POS_C(D)}(x)}{|U|}
\]

Where \(\tau_C(D)\) the significance of fuzzy attribute \(C\) with respect to fuzzy class label \(D\).

\(|U|\) Indicates the number of instances in the universe.

3.2 **Fuzzification**

3.2.1 **Fuzzification of Class Label Attribute**

- **Step 1**: For each class \(k \in \{1, 2, 3, \ldots, p\}\), calculating the center \(c_k\) of class \(k\);
- **Step 2**: For each instance \(x_i \in U\), calculating the distance \(d_{ik}\) between \(x_i\) and \(c_k\);
- **Step 3**: For each instance \(x_i \in U\), calculating the fuzzy membership degree of class as follows,

\[
\mu_k(x_i) = \frac{(d_{ik}^{-2})^{-1}}{\sum_{k=1}^{p} (d_{ik}^{-2})^{-1}} \quad (1 \leq i \leq N; 1 \leq k \leq p).
\]

3.2.2 **Fuzzification of real valued attributes**

**Input**: A Information System with real-valued attributes.

**Output**: A Fuzzy System with fuzzy attributes (The fuzzy class label attribute remains unchanged).

- **Step 1**: For \(i = 1 \rightarrow n\);
- **Step 2**: Let \(k = 2\) and \(T = T_0\);
- **Step 3**: Constructing the membership function of real valued attributes \(a_i\) as follows:

**Step 3.1**: Clustering the values \(\{a_{i1}, a_{i2}, \ldots, a_{iN}\}\) of \(a_i\) into \(k\) clusters, \(m_1, m_2, \ldots, m_k\) are the centers of the \(k\) clusters respectively.
Step 3.2: For \( t=1 \rightarrow k \)

If \((t=1)\) then

\[
\mu_{V_i}(a_j) =\begin{cases} 
\frac{a_i-m_i}{a_{\min}-m_i} \times 0.5 & \text{if } a_{\min} \leq a_i \leq m_1 \\
\frac{a_i-m_i}{m_i-m_i} & \text{if } a_i \leq m_2 \\
0 & \text{otherwise}
\end{cases}
\]

Else if \((t=k)\) then

\[
\mu_{V_i}(a_j) =\begin{cases} 
\frac{a_i-m_k}{m_{k-1}-m_k} & \text{if } m_{k-1} \leq a_i \leq m_k \\
\frac{a_i-m_k}{a_{\max}-m_k} \times 0.5 & \text{if } a_k \leq a_i \leq a_{\max} \\
0 & \text{otherwise}
\end{cases}
\]

Else

\[
\mu_{V_i}(a_j) =\begin{cases} 
\frac{a_i-m_j}{m_{j-1}-m_j} & \text{if } m_{j-1} \leq a_i \leq m_j \\
\frac{a_i-m_j}{m_{j-1}-m_j} & \text{if } m_i \leq a_i \leq m_{j+1} \\
0 & \text{otherwise}
\end{cases}
\]

Step 4: Calculating the increment \( \Delta \text{Gain}(a_i) \) of information gain of attribute \( a_i \);

Step 5: If \((k = 2 \text{ or } \Delta \text{Gain}(a_i) > T_0)\) then \( k = k + 1 \), and go to Step 3; Else exit and \( k = k - 1 \);

Step 6: If \((i < n)\) then \( i \leftarrow i + 1 \), and go to Step 2; Else exit.

3.3 Algorithm for Generating Fuzzy Decision Tree

Input: Fuzzy System = \( \{ U, C \cup D, V, f \} \)

Output: A group of fuzzy classification rules extracted from the generated fuzzy decision tree \( T \).

Step 1: Preparing the Fuzzy System.

Step 2: Selecting the expanded attribute.

Step 2.1: For each fuzzy attribute \( C_i \) and its fuzzy linguistic term \( C_{i_i} \) \((1 \leq i \leq n)\), the significance of \( C_i \) with respect to the fuzzy class label attribute \( D \) is calculated by using formula (8).

Step 2.2: Selecting \( i_0 \) according to \( i_0 = \text{Argmax}_{1 \leq i \leq n} \{ \tau_{A_i}(C) \} \) \((C_{i_0} \text{ is the expanded attribute})\).

Step 3: If the terminal conditions cannot be satisfied, then partition \( U \), and recursively select the expanded attribute until a fuzzy decision tree is generated.

Step 4: Extracting fuzzy classification rules from the fuzzy decision tree \( T \).

3.4 Terminal Conditions

Terminal conditions used for generating a fuzzy decision tree are as follows

i) If the classification truth degree of a branch with respect to one class exceeds a given threshold \( \beta \), then the branch is terminated as a leaf node.

ii) At a branch, if no attribute can be expanded, the branch is terminated as a leaf node or as null node.

iii) If one of the conditions mentioned above is satisfied, the expanding of the branch in a fuzzy decision tree will be terminated.

Suppose \( P \) and \( Q \) are two fuzzy sets on \( U \). The truth degree of the fuzzy set \( P \) with respect to the fuzzy set \( Q \) is defined as follows

\[
\beta_T(P, Q) = \frac{\tau_P(Q)}{\sum_{x \in U} \mu_P(x)} \tag{10}
\]

4 Illustrating Fuzzy Rough Decision Tree Concept with Example
In this section, we will demonstrate the process of generating fuzzy decision tree based on fuzzy rough method by using the small Fuzzy System with the 9 instances in the Table 2. The universe \( U = \{O_1, O_2, O_3, \ldots, O_9\} \). The four fuzzy attributes are sepal length, sepal width, petal length and petal width. The fuzzy class label attribute is Type.

Let the threshold of classification truth degree \( \beta = 0.78 \). By calculating the significance of each fuzzy attribute with respect to the fuzzy class label attribute \( D = \{\text{Type}\} \), we have

\[
\tau_{\text{sepallength}}(\text{Type}) = \frac{\sum_{O \in U} \mu_{\text{POS}_{\text{sepallength}}(\text{Type})}(O)}{|U|} = 0.31122
\]

\[
\tau_{\text{sepalwidth}}(\text{Type}) = \frac{\sum_{O \in U} \mu_{\text{POS}_{\text{sepalwidth}}(\text{Type})}(O)}{|U|} = 0.1768
\]

\[
\tau_{\text{petallength}}(\text{Type}) = \frac{\sum_{O \in U} \mu_{\text{POS}_{\text{petallength}}(\text{Type})}(O)}{|U|} = 0.3504
\]

\[
\tau_{\text{sepallength}}(\text{Type}) = \frac{\sum_{O \in U} \mu_{\text{POS}_{\text{sepallength}}(\text{Type})}(O)}{|U|} = 0.7291
\]

Since, the fuzzy attribute petal width has the biggest significance; it is selected as the root node. There are three branches (short, medium, and long) from the root node petal width.

Calculating the value of classification truth degree by using the formulae (10)

\[
\tau_{\text{short-petalwidth}}(\text{setosa}) = (0.9474 + 0.9474 + 0.9474 + 0.0526) = 2.8948
\]

\[
\tau_{\text{medium-petalwidth}}(\text{setosa}) = (0.9474 + 0.9474 + 0.9474 + 0.0526) = 2.8948
\]

\[
\tau_{\text{long-petalwidth}}(\text{setosa}) = (0.9474 + 0.9474 + 0.9474 + 0.0526) = 2.8948
\]

\[
\tau_{\text{short-petalwidth}}(\text{setosa}) = (0.9474 + 0.9474 + 0.9474 + 0.0526) = 2.8948
\]

So, \( T(P, Q) = 2.8948 / 3.0526 = 0.9483 \) which meets the terminal condition (1) and becomes a leaf node with label iris-setosa.

At the branch medium,

\( T(\text{medium}, \text{setosa}) < \beta \)

\( T(\text{medium}, \text{versicolor}) < \beta \)

\( T(\text{medium}, \text{virginica}) < \beta \)

So need to consider additional attributes and we have three fuzzy partitions

1) Medium-Sepal length = \{medium \( \cap C_{1i}\) \( i=1, 2 \)

Where \( \{C_{11}, C_{12}, C_{13}\} = \{\text{short, long}\} \)

2) Medium-Sepal width = \{medium \( \cap C_{2i}\) \( i=1, 2 \)

3) Medium- Petal length = \{ Medium \( \cap C_{3i}\) \( i=1, 2 \)

Where \( \{C_{31}, C_{32}\} = \{\text{short, long}\} \)

Similarly, calculating the \( T_{\text{medium-sepalwidth}}(\text{Type}) \), \( T_{\text{medium-sepalwidth}}(\text{Type}) \) and \( T_{\text{medium-petallength}}(\text{Type}) \) then, select the node that has biggest value and add the node to the branch medium in the fuzzy decision tree and repeat it till the terminal conditions are met.

Finally generated fuzzy decision tree look like this

![Decision Tree Diagram](image-url)
5 Extracting Classification Rules from the Fuzzy Decision Tree

In this step, consider each path of the fuzzy decision tree from root node to leaf node converted in to a fuzzy classification rule with the highest classification truth degree.

For the above fuzzy decision tree, five rules are obtained.

**Rule 1:** If Petal width is short, then Class is Setosa (0.94823);

**Rule 2:** If Petal width is Medium and Sepal length is Short, then Class is Setosa (1);

**Rule 3:** If Petal width is Medium and Sepal length is Long, then Class is Versicolor (1);

**Rule 4:** If Petal width is long and Sepal Width is short, then Class is Virginica (0.935);

**Rule 5:** If Petal width is long and Sepal Width is long, then Class is Virginica (0.834);

6 Results and Experimental Analysis

The effectiveness of fuzzy rough method is tested through numerical experiments in the environment of MATLAB 7.0 on a core 2 duo PC. Totally our experiments conducted on 6 datasets are from UCI. The UCI datasets are Iris Dataset (DS1), Pima Dataset (DS2), Ionosphere Dataset (DS3), Breast Cancer Dataset-WDBC (DS4), Glass Dataset (DS5), and Blood Transfusion Service Center Dataset (DS6).

<table>
<thead>
<tr>
<th>DB</th>
<th>Fuzzy ID3 (Average accuracy)</th>
<th>Our proposed method (Average accuracy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>0.9533</td>
<td>0.9333</td>
</tr>
<tr>
<td>DS2</td>
<td>0.6948</td>
<td>0.8831</td>
</tr>
<tr>
<td>DS3</td>
<td>0.6556</td>
<td>0.9139</td>
</tr>
<tr>
<td>DS4</td>
<td>0.6843</td>
<td>0.6686</td>
</tr>
<tr>
<td>DS5</td>
<td>0.8220</td>
<td>0.8273</td>
</tr>
<tr>
<td>DS6</td>
<td>0.6307</td>
<td>0.6600</td>
</tr>
</tbody>
</table>

6 Conclusions

There are several criterions to select expanded attribute for generating fuzzy decision trees. Fuzzy classification entropy and classification ambiguity are used till now. However, in this paper, which uses the significance of fuzzy conditional attribute with respect to the fuzzy decision attribute as a measure to select the expanded node to generate fuzzy decision tree.

The fuzzy rough algorithm outperforms Fuzzy ID3 on several standard testing datasets. The classification average accuracy is significantly better than in case of our method compared to that of Fuzzy ID3.

References


