Fuzzy Rough Sets with Hierarchical Quantitative Attributes

Tzung-Pei Hong1*, Yan-Liang Liou2 and Shyue-Liang Wang3
1Department of Computer Science and Information Engineering
2Department of Electrical Engineering
3Department of Information Management
National University of Kaohsiung
Kaohsiung, 811, Taiwan, R.O.C.
tphong@nuk.edu.tw, m0945108@mail.nuk.edu.tw, slwang@nuk.edu.tw

Abstract

Machine learning can extract desired knowledge and ease the development bottleneck in building expert systems. Among the proposed approaches, deriving classification rules from training examples is the most common. Given a set of examples, a learning program tries to induce rules that describe each class. The rough-set theory has served as a good mathematical tool for dealing with data classification problems. It adopts the concept of equivalence classes to partition training instances according to some criteria. In the past, we thus proposed a fuzzy-rough approach to produce a set of certain and possible rules from quantitative data. Attributes are, however, usually organized into hierarchy in real applications. This paper thus extends our previous approach to deal with the problem of producing a set of cross-level maximally general fuzzy certain and possible rules from examples with hierarchical and quantitative attributes. The proposed approach combines the rough-set theory and the fuzzy-set theory to learn. It is more complex than learning from single-level values, but may derive more general knowledge from data. Fuzzy boundary approximations, instead of upper approximations, are used to find possible rules, thus reducing some subsumption checking. Some pruning heuristics are adopted in the proposed algorithm to avoid unnecessary search. A simple example is also given to illustrate the proposed approach.

Keywords: machine learning, rough set, certain rule, possible rule, hierarchical value, quantitative value.

*Corresponding author. Also at Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan.
1. Introduction

Expert systems have been widely used in domains where mathematical models cannot easily be built, human experts are not available or the cost of querying an expert is high. Although a wide variety of expert systems have been built, knowledge acquisition remains a development bottleneck. Usually, a knowledge engineer is needed to establish a dialog with a human expert and to encode the knowledge elicited into a knowledge base to produce an expert system. The process is, however, very time-consuming [1][17]. Building a large-scale expert system involves creating and extending a large knowledge base over the course of many months or years. Shortening the development time is thus the most important factor for the success of an expert system. Machine-learning techniques have thus been developed to ease the knowledge-acquisition bottleneck. Among the proposed approaches, deriving rules from training examples is the most common [7][11][12]. Given a set of examples, a learning program tries to induce rules that describe each class.

Recently, the rough-set theory has been used in reasoning and knowledge acquisition for expert systems [2][5][14]. It was proposed by Pawlak in 1982 [15][16] with the concept of equivalence classes as its basic principle. Several applications and extensions of the rough-set theory have also been proposed. Examples are Orlowska's reasoning with incomplete information [14], Germano and Alexandre's knowledge-base reduction [4], Lingras and Yao's data mining [10]. Lambert-Torres et al. found unimportant attributes from lower and upper approximations and deleted them from a database [8]. Zhong et al. proposed a new incremental learning algorithm based on the generalization distribution table, which maintained the probabilistic relationships between the possible instances and the possible concepts [23]. Yao formed a stratified granulation structure with respect to different levels of rough set approximations by incrementally clustering objects with the same characteristics together [19]. Also, Lee et al. simplified classification rules for data mining
using rough set theory [9]. Tsumoto presented a knowledge discovery system based on rough sets and attribute-oriented generalization [18]. It was used not only to acquire several sets of attributes important for classification, but also to evaluate how precisely the attributes of a database were able to classify data. Many researches about this field are still in progress [21][22].

Training data in real-world applications sometimes consist of quantitative values. Fuzzy-set concepts are often used to represent quantitative data expressed in linguistic terms and membership functions in intelligent systems because of its simplicity and similarity to human reasoning. Dubois and Prade combined rough sets and fuzzy sets together in order to get a more accurate account of imperfect information [3]. They built up a very good theoretic basis for fuzzy rough sets. Also, Nakamura predefined similarity matrices and used fuzzy rough sets to logic reasoning [13]. In the past, we also proposed a method which combined rough-set theory and fuzzy-set theory to deal with the problem of producing a set of certain and possible rules from quantitative data [6].

Attributes are usually organized into hierarchy in real applications. Deriving rules on multiple concept levels may thus lead to the discovery of more general and important knowledge from data. It is, however, more complex than learning rules from training examples with single-level values. This paper thus extends our previous approach to deal with the problem of producing a set of cross-level maximally general fuzzy certain and possible rules from examples with hierarchical and quantitative attributes. Fuzzy boundary approximations, instead of upper approximations, are used to find possible rules, thus reducing some subsumption checking. Some pruning heuristics are adopted in the proposed algorithm to avoid unnecessary search. Rule effectiveness for future data is also derived from these membership values.

The remainder of this paper is organized as follows. The rough-set theory and fuzzy-set
concept are reviewed in Sections 2 and 3. Management of hierarchical attribute values by rough sets is described in Section 4. The notation and definitions used in this paper are given in Section 5. A new learning algorithm which can process hierarchical and quantitative attributes by fuzzy rough sets is proposed in Section 6. An example to illustrate the proposed algorithm is given in Section 7. Some discussion is taken in Section 8. Conclusions and future works are finally given in Section 9.

2. Review of related works.

2.1 The rough-set theory

The rough-set theory, proposed by Pawlak in 1982 [15],[16], can serve as a new mathematical tool for dealing with data classification problems. It adopts the concept of equivalence classes to partition training instances according to some criteria. Two kinds of partitions are formed in the mining process: lower approximations and upper approximations, from which certain and possible rules can easily be derived.

Formally, let $U$ be a set of training examples (objects), $A$ be a set of attributes describing the examples, $C$ be a set of classes, and $V_j$ be a value domain of an attribute $A_j$. Also let $v^{(i)}_j$ be the value of attribute $A_j$ for the $i$-th object $Obj^{(i)}$. When two objects $Obj^{(i)}$ and $Obj^{(k)}$ have the same value of attribute $A_j$, (that is, $v^{(i)}_j = v^{(k)}_j$), $Obj^{(i)}$ and $Obj^{(k)}$ are said to have an indiscernibility relation (or an equivalence relation) on attribute $A_j$. Also, if $Obj^{(i)}$ and $Obj^{(k)}$ have the same values for each attribute in subset $B$ of $A$, $Obj^{(i)}$ and $Obj^{(k)}$ are also said to have an indiscernibility (equivalence) relation on attribute set $B$. These equivalence relations thus partition the object set $U$ into disjoint subsets, denoted by $U/B$, and the partition including $Obj^{(i)}$ is denoted $B(Obj^{(i)})$. The sets of equivalence classes for subset $B$ are referred to as $B$-elementary sets.
The rough-set approach analyzes data according to two basic concepts, namely the lower and the upper approximations of a set. Let $X$ be an arbitrary subset of the universe $U$, and $B$ be an arbitrary subset of attribute set $A$. The lower and the upper approximations for $B$ on $X$, denoted $B^-(X)$ and $B^+(X)$ respectively, are defined as follows:

$$B^-(X) = \{ x \mid x \in U, B(x) \subseteq X \}, \text{ and}$$
$$B^+(X) = \{ x \mid x \in U \text{ and } B(x) \cap X \neq \emptyset \}.$$

Elements in $B^-(X)$ can be classified as members of set $X$ with full certainty using attribute set $B$, so $B^-(X)$ is called the lower approximation of $X$. Similarly, elements in $B^+(X)$ can be classified as members of the set $X$ with only partial certainty using attribute set $B$, so $B^+(X)$ is called the upper approximation of $X$. After the lower and the boundary approximations have been found, the rough-set theory can then be used to derive both certain and uncertain information and induce certain and possible rules from them.

### 2.2 The fuzzy-set concepts

The fuzzy-set theory was first proposed by Zadeh in 1965 [20]. It is primarily concerned with quantifying and reasoning using natural language in which words can have ambiguous meanings [20][24][25]. This can be thought of as an extension of traditional crisp sets in which each element must either be in or not in a set.

Formally, the process by which individuals from a universal set $X$ are determined to be either members or non-members of a crisp set can be defined by a characteristic or discrimination function [20]. For a given crisp set $A$, this function assigns a value $\mu_A(x)$ to every $x \in X$ such that:

$$\mu_A(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{if and only if } x \notin A. \end{cases}$$
This kind of function can be generalized such that the values assigned to the elements of the universal set fall within specified ranges, referred to as the membership grades of these elements in the set, with larger values denoting higher degrees of set membership. Such a function is called the membership function, $\mu_A(x)$, by which a fuzzy set $A$ is usually defined. This function is represented by:

$$\mu_A : X \rightarrow [0,1],$$

where $[0, 1]$ denotes the interval of real numbers from 0 to 1, inclusive. The function can also be generalized to any real interval and is not restricted to $[0,1]$.

A special notation is often used in the literature to represent fuzzy sets. Assume that $x_1$ to $x_n$ are the elements in fuzzy set $A$, and $\mu_1$ to $\mu_n$ are, respectively, their grades of membership in $A$. $A$ is then usually represented as follows:

$$A = \mu_1 / x_1 + \mu_2 / x_2 + \ldots + \mu_n / x_n.$$  

An $\alpha$-cut of a fuzzy set $A$ is a crisp set $A_\alpha$ that contains all elements in the universal set $X$ with membership grades in $A$ greater than or equal to a specified value of $\alpha$. This definition can be written as:

$$A_\alpha = \{ x \in X \mid \mu_A(x) \geq \alpha \}.$$  

The scalar cardinality of a fuzzy set $A$ defined on a finite universal set $X$ is the summation of the membership grades of all the elements of $X$ in $A$. Thus,

$$|A| = \sum_{x \in X} \mu_A(x).$$  

Among operations on fuzzy sets are the basic and commonly used complementation, union and intersection, as proposed by Zadeh. They are defined as follows.

1. The complementation of a fuzzy set $A$ is denoted by $\neg A$, and the membership function of $\neg A$ is given by:
\( \mu_{-A}(x) = 1 - \mu_{A}(x), \forall x \in X. \)

(2) The intersection of two fuzzy sets \( A \) and \( B \) is denoted by \( A \cap B \), and the membership function of \( A \cap B \) is given by:
\[
\mu_{A \cap B}(x) = \min \{ \mu_A(x), \mu_B(x) \}, \forall x \in X.
\]

(3) The union of fuzzy sets \( A \) and \( B \) is denoted by \( A \cup B \), and the membership function of \( A \cup B \) is given by:
\[
\mu_{A \cup B}(x) = \max \{ \mu_A(x), \mu_B(x) \}, \forall x \in X.
\]

The above fuzzy operations are used in the proposed learning algorithm to find linguistic certain and possible rules.

### 2.3. Hierarchical Attributes

Most of the previous studies on rough sets focused on finding certain rules and possible rules on the single concept level. However, hierarchical attributes are usually predefined in real-world applications and can be represented by hierarchy trees. Terminal nodes on the trees represent actual attribute values appearing in training examples; internal nodes represent value clusters formed from their lower-level nodes. Deriving rules on multiple concept levels may lead to the discovery of more general and important knowledge from data. A simple example for attribute *Transport* is given in Figure 1.
Figure 1: An example of predefined hierarchical values for attribute *Transport*

In Figure 1, the attribute *Transport* falls into two general categories: *Train* and *Car*. *Train* can be further classified into two more specific categories *Express Train* and *Ordinary Train*. Similarly, assume *Car* is divided into *Expensive Car* and *Cheap Car*. Only the terminal attribute nodes (*Express Train, Ordinary Train, Expensive Car, Cheap Car*) can appear in training examples.

The concept of equivalence classes in the rough set theory makes it very suitable for finding cross-level certain and possible rules from training examples with hierarchical values. The equivalence class of a non-terminal-level attribute value for attribute \(A_j\) can be easily found by the union of its underlying terminal-level equivalence classes for \(A_j\). Also, the equivalence class of a cross-level attribute value combination for more than two attributes can be derived from the intersection of the equivalence classes of its single attribute values. In this paper, we will propose a fuzzy-rough learning algorithm for deriving cross-level certain and possible rules from training examples with hierarchical quantitative attribute values.

4. Fuzzy-Rough Sets

In the past, we proposed a method which combined rough-set theory and fuzzy-set theory to deal with the classification problem [6]. It is extended here to find a set of cross-level maximally general fuzzy certain and possible rules from examples with hierarchical and quantitative attributes. Some definitions about fuzzy approximations are introduced below.

When the same linguistic term \(R_{jk}\) of an attribute \(A_j\) exists in two fuzzy objects \(Obj^{(i)}\) and
with membership values $f_{jk}^{(i)}$ and $f_{jk}^{(r)}$ equal to or larger than a certain $\alpha$ value, $Obj^{(i)}$ and $Obj^{(r)}$ are said to have a fuzzy $\alpha$-indiscernibility relation (or fuzzy $\alpha$-equivalence relation) on attribute $A_j$ with membership value $\min(f_{jk}^{(i)} \cap f_{jk}^{(r)})$. Also, if the same linguistic terms of an attribute subset $B$ exist in both $Obj^{(i)}$ and $Obj^{(r)}$ with membership values equal to or larger than $\alpha$, $Obj^{(i)}$ and $Obj^{(r)}$ are said to have a fuzzy $\alpha$-indiscernibility relation (or a fuzzy $\alpha$-equivalence relation) on attribute subset $B$ with a membership value equal to the minimum of all the membership values. These fuzzy $\alpha$-equivalence relations thus partition the fuzzy object set $U$ into several fuzzy subsets that may overlap, and the result is denoted by $U/B$. The set of partitions, based on $B$ and including $Obj^{(i)}$, is denoted $B(Obj^{(i)})$. Thus, $B(Obj^{(i)})=\{(B_1(Obj^{(i)}), \mu_{B_1}(Obj^{(i)})), \ldots, (B_r(Obj^{(i)}), \mu_{B_r}(Obj^{(i)}))\}$, where $r$ is the number of partitions included in $B(Obj^{(i)})$, $B_j(Obj^{(i)})$ is the $j$-th partition in $B(Obj^{(i)})$, and $\mu_{B_j}(Obj^{(i)})$ is the membership value of the $j$-th partition.

Example 1: Consider the following three fuzzy objects shown in Table 1. $Obj^{(1)}$ has a normal systolic pressure with a membership value of 0.1 and a high systolic pressure with a membership value of 0.75. $Obj^{(1)}$ has also a normal diastolic pressure with a membership value of 0.4 and a high diastolic pressure with a membership value of 0.8. Furthermore, $Obj^{(1)}$ is classified as having a high blood pressure. $Obj^{(2)}$ and $Obj^{(3)}$ can be explained in a similar way.

<table>
<thead>
<tr>
<th>Object</th>
<th>Systolic Pressure (SP)</th>
<th>Diastolic Pressure (DP)</th>
<th>Blood Pressure (BP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Obj^{(1)}$</td>
<td>(0.1/N + 0.75/H)</td>
<td>(0.4/N + 0.8/H)</td>
<td>H</td>
</tr>
<tr>
<td>$Obj^{(2)}$</td>
<td>(1/H)</td>
<td>(0.16/N + 0.6/H)</td>
<td>H</td>
</tr>
<tr>
<td>$Obj^{(3)}$</td>
<td>(0.5/L + 0.3/N)</td>
<td>(0.4/N + 0.3/L)</td>
<td>L</td>
</tr>
</tbody>
</table>
Let the $\alpha$ value be set at 0.1. Since the same linguistic term ($N$) for attribute $SP$ exists in both $Obj^{(1)}$ and $Obj^{(3)}$ and both the membership values are equal to or larger than 0.1, they have a fuzzy indiscernibility relation on the fuzzy term $SP.N$ and thus form a fuzzy equivalence class with a membership value of $\min(0.1, 0.3)$. The other fuzzy indiscernibility relations can be similarly derived. $U\{SP\}$ has thus been formed and may be represented as follows:

$$U\{SP\} = \{(\{Obj^{(1)}, Obj^{(3)}\}, 0.1) \,(\{Obj^{(1)}, Obj^{(2)}\}, 0.75) \,(\{Obj^{(3)}\}, 0.5)\}.$$ 

Similarly,

$$U\{DP\} = \{(\{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}\}, 0.16) \,(\{Obj^{(1)}, Obj^{(2)}\}, 0.6) \,(\{Obj^{(3)}\}, 0.3)\}.$$ 

Also, $SP(Obj^{(1)}) = \{(\{Obj^{(1)}, Obj^{(3)}\}, 0.1) \,(\{Obj^{(1)}, Obj^{(2)}\}, 0.75)\}$. It can be easily seen that $Obj^{(1)}$ exists in more than one fuzzy equivalence class. The set of fuzzy equivalence classes for a subset set $B$ is referred to as a fuzzy $B$-elementary set.

Fuzzy $\alpha$-lower and fuzzy $\alpha$-boundary approximations are defined below. Let $X$ be an arbitrary subset of the universe $U$, and $B$ be an arbitrary subset of the attribute set $A$. The fuzzy lower and the fuzzy boundary approximations under the threshold value $\alpha$ for $B$ on $X$, denoted $B^\ast(X)$ and $B^\ast(X)$, respectively, are defined as follows:

$$B^\ast(X) = \{(B_k(x), \mu_{B_k}(x)) \,| \, x \in U, B_k(x) \subseteq X, 1 \leq k \leq |B(x)|\},$$

and

$$B^\ast(X) = \{(B_k(x), \mu_{B_k}(x)) \,| \, x \in U, B(x) \cap X \neq \emptyset \,\text{and} \, B(x) \subset X, 1 \leq k \leq |B(x)|\}.$$ 

Elements in $B^\ast(x)$ can be classified as members of set $X$ with full certainty using attribute set $B$. Also, their membership values may be considered effectiveness measures of fuzzy lower approximations for future data. A low membership value with a fuzzy lower approximation means the lower approximation will have a low tolerance (or effectiveness) on future data. In this case, the fuzzy lower-approximation partitions have a high probability
of being removed when future data are considered. All of the partitions are, however, valid for the current data set and can be used to correctly classify its elements.

On the other hand, elements in $B^*(x)$ can be classified as members of set $X$ with only partial certainty using attribute set $B$, and their certainty degrees can be calculated from the membership values of elements in the boundary approximations.

**Example 2:** Continuing from Example 1, assume $X = \{Obj^{j1}, Obj^{j2}\}$. The fuzzy lower approximation and the fuzzy boundary approximation for attribute $SP$ according to $X$ can be calculated as follows:

$$SP_\ast(X) = (\{Obj^{j1}, Obj^{j2}\}, 0.75), \text{ and}$$

$$SP^\ast(X) = \{\{Obj^{j1}, Obj^{j3}\}, 0.1\}.$$

After the fuzzy lower and the fuzzy boundary approximations have been found, certain and uncertain information can be analyzed, and rules can then be derived.

### 5. The Proposed Algorithm

In the section, a learning algorithm based on fuzzy rough sets is proposed to find fuzzy cross-level certain and possible rules from training data with hierarchical and quantitative attribute values. According to the definitions of the fuzzy lower approximation and the fuzzy upper approximation, it is easily seen that the fuzzy upper approximation includes the fuzzy lower approximation. Thus each fuzzy certain rule derived from the fuzzy lower approximation will also be derived from the fuzzy upper approximation. It thus causes redundant derivation and wastes computational time. The proposed algorithm thus uses the fuzzy boundary approximation, instead of the fuzzy upper approximation, to derive the pure fuzzy possible rules. It can thus reduce the subsumption checking needed. For convenience, the symbol $B^\ast(X)$ is used from here on to represent the fuzzy boundary approximation of
attribute subset $B$ on $X$, instead of the fuzzy upper approximation.

The proposed algorithm first transforms each quantitative value into a fuzzy set of linguistic terms using membership functions and finds the terminal-level fuzzy elementary sets of single attributes. These fuzzy equivalence classes can then be used later to find the non-terminal-level fuzzy elementary sets of single attributes and the cross-level fuzzy elementary sets of attribute combinations. Fuzzy lower approximations are used to derive fuzzy certain rules. Fuzzy boundary approximations, as mentioned above, are used to find fuzzy possible rules. The algorithm calculates the fuzzy lower and the fuzzy boundary approximations of single attributes from the terminal level to the higher level. After that, the fuzzy lower and the fuzzy boundary approximations of more than one attribute are derived based on the results of single attributes. Some pruning heuristics are also used to avoid unnecessary search. The rule-derivation process based on these approximations is then performed to find maximally generally fuzzy certain rules and fuzzy possible rules. The details of the proposed learning algorithm are described as follows.

**A learning algorithm processing hierarchical and quantitative attributes by fuzzy rough sets:**

Input: A quantitative data set with $n$ objects, $m$ hierarchical attributes, a set of membership functions, and a value $\alpha$ for $\alpha$-cut.

Output: A set of cross-level certain and possible rules.

Step 1: Partition the object set into disjoint subsets according to class labels. Denote each set of objects belonging to the same class $C_i$ as $X_i$.

Step 2: Transform the quantitative value $v_{ij}^{(i)}$ of each object $Obj^{(i)}$, $i = 1$ to $n$, for each appearing terminal-level node $A_j$ in $Obj^{(i)}$ into a fuzzy set $f_{ij}^{(i)}$, represented as:
\[
\left( \frac{f_{jk}^{(i)}}{R_{j1}} \right) + \frac{f_{jk}^{(i)}}{R_{j2}} + \ldots + \frac{f_{jk}^{(i)}}{R_{jk}} \right),
\]

using the given membership functions, where \( R_{jk} \) is the \( k \)-th fuzzy region of a terminal-level attribute node \( A_j \), \( f_{jk}^{(i)} \) is \( v^{(i)} \)'s fuzzy membership value in region \( R_{jk} \), and \( l (=|A_j|) \) is the number of fuzzy regions for \( A_j \).

Step 3: Remove the linguistic term \( R_{jk} \) from a fuzzy set \( f_j^{(i)} \) with its \( f_{jk}^{(i)} < \alpha \).

Step 4: Find the terminal-level fuzzy elementary sets of single attributes using fuzzy operations based on the definitions in Section 4.

Step 5: Set \( l = 1 \), where \( l \) is used to count the number of classes currently being processed.

Step 6: Compute the fuzzy lower approximation of each single terminal-level attribute node \( A_j^t \) for class \( X_i \) as:

\[
A_j^t(X_i) = \{ (A_{jk}^t(x), \mu_{A_{jk}^t}(x)) \mid x \in U, \ A_{jk}^t(x) \subseteq X_i, \ 1 \leq k \leq |A_j^t(x)| \}
\]

where \( A_{jk}^t(x) \) is the terminal-level fuzzy equivalence class including object \( x \) and derived from the \( k \)-th fuzzy region of attribute node \( A_j^t \), and \( |A_j^t(x)| \) is the number of fuzzy regions for \( A_j^t \).

Step 7: Compute the fuzzy boundary approximation of each single terminal-level attribute node \( A_j^t \) for class \( X_i \) as:

\[
A_j^w(X_i) = \{ (A_{jk}^t(x), \mu_{A_{jk}^t}(x)) \mid x \in U, \ A_{jk}^t(x) \cap X_i \neq \phi, A_{jk}^t(x) \not\subseteq X_i, 1 \leq k \leq |A_j^t(x)| \}
\]

Step 8: Compute the fuzzy lower and fuzzy boundary approximations of each single non-terminal-level attribute node \( A_j'' \) for class \( X_i \) from the terminal level to the root level by the following substeps:
(a) Derive the fuzzy equivalence class $A_{j_{nt}}^{nt}$ of the $i$-th fuzzy region for the non-terminal-level attribute node $A_{j}^{nt}$ on level $k$ by the union of the equivalence classes of the same fuzzy regions in its underlying nodes.

(b) Put the equivalence class of $A_{j_{nt}}^{nt}$ into the $k$-level lower approximation for non-terminal-level attribute node $A_{j}^{nt}$ if all its underlying equivalence classes of the same fuzzy regions are in the $(k+1)$-level lower approximation for attribute node $A_{j}^{nt}$.

(c) Put the equivalence class of $A_{j_{nt}}^{nt}$ in the $k$-level boundary approximation for attribute node $A_{j}^{nt}$ if at least one of its underlying equivalence classes of the same fuzzy regions is in the $(k+1)$-level lower or boundary approximation for attribute node $A_{j}^{nt}$.

Step 9: Set $q = 2$, where $q$ is used to count the number of attributes currently being processed.

Step 10: Compute the fuzzy lower and the fuzzy boundary approximations of each attribute set $B_j$ with $q$ attributes (on any levels) for class $X_l$ from the terminal level to the root level by the following substeps. Only a fuzzy region from an attribute can be put in the combination.

(a) Skip the combinations which have the equivalence class of at least one of its subsets already in the lower approximation for $X_l$.

(b) Derive the equivalence class of each remaining combination by the intersection of the equivalence classes of its corresponding single attribute regions. Set the membership values for the derived equivalence class as the minimum of their membership values.
(c) Put the equivalence class \( B_j(x) \) of each combination in substep (b) into the lower approximation class \( X_l \) if \( B_j(x) \subseteq X_l \).

(d) Put the equivalence class \( B_j(x) \) of each combination in substep (b) into the boundary approximation class if \( B_j(x) \cap X_l \neq \emptyset \) and \( B_j(x) \nsubseteq X_l \).

Step 11: Set \( q = q + 1 \) and repeat Step 10 until \( q > m \).

Step 12: Derive the certain rules from the fuzzy lower approximation \( B^*(X_l) \) of any subset \( B \), and set the membership values of elements in the lower approximation as effectiveness measures for future data.

Step 13: Remove certain rules more specific than others and keep the more general ones.

Step 14: Derive the possible rules from the fuzzy upper approximation \( B^*(X_l) \) of any subset \( B \), set the membership values of elements in the upper approximation as effectiveness measures for future data, and calculate the plausibility measure of each rule for \( B_k(x) \) as:

\[
p(B_k(x)) = \frac{\sum_{x \in B_k(x) \cap X_l} \mu_{B_k}(x)}{\sum_{x \in B_k(x)} \mu_{B_k}(x)}.
\]

Step 15: Remove possible rules with condition parts more specific and plausibility values equal to or smaller than those of some other possible or certain rules.

Step 16: Set \( l = l + 1 \) and repeat Steps 6 to 15 until \( l > c \).

Step 17: Output the fuzzy certain rules and fuzzy possible rules.

After Step 17, certain and possible rules can then be derived, and can serve as meta-knowledge concerning the given data set.
6. An Example

In this section, a simple example is given to show how the proposed algorithm can be used to generate linguistic certain and possible rules from quantitative training data with hierarchical values. There are two decision attributes \( A = \{\text{Transport, Residence}\} \), and a class attribute \( C = \{\text{Consumption Style}\} \). Both the attributes have a simple hierarchy as shown in Figures 2 and 3. There are two levels of hierarchical attribute values for attributes \textit{Transport} and \textit{Residence}. The roots representing the generic names of attributes are located on level 0 (such as “Transport” and “Residence”), the terminal nodes representing actual values (such as “Imported car”) are on level 1. Only values of terminal nodes can appear in training examples. Assume the class has only two possible values: \{High (H), Low (L)\}.

![Figure 2: Hierarchy of Transport](image)

![Figure 3: Hierarchy of Residence](image)

Assume the training data set is shown in Table 2.

<p>| Table 2: The training data used in this example |</p>
<table>
<thead>
<tr>
<th>Object</th>
<th>Transport</th>
<th>Residence</th>
<th>Consumption Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obj^{(1)}</td>
<td>Imported car:190</td>
<td>House:1700</td>
<td>High</td>
</tr>
<tr>
<td>Obj^{(2)}</td>
<td>Imported car:210</td>
<td>House:1800</td>
<td>High</td>
</tr>
<tr>
<td>Obj^{(3)}</td>
<td>Native car:50</td>
<td>Building:230</td>
<td>Low</td>
</tr>
<tr>
<td>Obj^{(4)}</td>
<td>Native car:42</td>
<td>Building:220</td>
<td>Low</td>
</tr>
<tr>
<td>Obj^{(5)}</td>
<td>Native car:45</td>
<td>House:350</td>
<td>Low</td>
</tr>
<tr>
<td>Obj^{(6)}</td>
<td>Native car:40</td>
<td>House:330</td>
<td>Low</td>
</tr>
<tr>
<td>Obj^{(7)}</td>
<td>Native car:145</td>
<td>Building:1300</td>
<td>High</td>
</tr>
<tr>
<td>Obj^{(8)}</td>
<td>Native car:130</td>
<td>Building:1400</td>
<td>High</td>
</tr>
<tr>
<td>Obj^{(9)}</td>
<td>Imported car:50</td>
<td>Building:250</td>
<td>Low</td>
</tr>
<tr>
<td>Obj^{(10)}</td>
<td>Imported car:60</td>
<td>Building:270</td>
<td>Low</td>
</tr>
</tbody>
</table>

Table 2 contains ten objects $U$={$Obj^{(1)}$, $Obj^{(2)}$, …, $Obj^{(10)}$}. Each training example has a quantity representing the cost (unit: ten thousands NT dollars) for the attribute value. Assume the membership functions of each terminal attribute node are shown in the Figure 4.
The proposed algorithm then processes the data in Table 2 as follows.

Step 1: Since two classes exist in the data set, two partitions are found as follows:

\[ X_H = \{Obj^{(1)}, Obj^{(2)}, Obj^{(6)}, Obj^{(7)}\}, \text{ and} \]
\[ X_L = \{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(8)}, Obj^{(9)}, Obj^{(10)}\}. \]

Step 2: The quantitative values of each object are transformed into fuzzy sets. Take the attribute *Transport* in *Obj*\(^{(1)}\) as an example. The value “190” is converted into the fuzzy set (0.1/IC.M + 0.58/IC.H) using the given membership functions, where “IC” is an abbreviation of “Imported Car”. Similarly, NC, HO and BU represent Native Car, House and Building, respectively. The converted results for all the objects are shown in Table 3.

<table>
<thead>
<tr>
<th>Object</th>
<th><em>Transport</em></th>
<th><em>Residence</em></th>
<th><em>Consumption Style</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Obj</em>(^{(1)})</td>
<td>0.1/IC.M+0.58/IC.H</td>
<td>0.21/HO.M+0.75/HO.H</td>
<td>High</td>
</tr>
<tr>
<td><em>Obj</em>(^{(2)})</td>
<td>1/IC.H</td>
<td>0.14/HO.M+0.83/HO.H</td>
<td>High</td>
</tr>
<tr>
<td><em>Obj</em>(^{(3)})</td>
<td>0.5/NC.L+0.1/NC.M</td>
<td>0.85/BU.L+0.1/BU.M</td>
<td>Low</td>
</tr>
<tr>
<td><em>Obj</em>(^{(4)})</td>
<td>0.9/NC.L+0.05/NC.M</td>
<td>0.9/BU.L+0.06/BU.M</td>
<td>Low</td>
</tr>
<tr>
<td><em>Obj</em>(^{(5)})</td>
<td>0.75/NC.L+0.1/NC.M</td>
<td>0.75/HO.L+0.2/HO.M</td>
<td>Low</td>
</tr>
<tr>
<td><em>Obj</em>(^{(6)})</td>
<td>1/NC.L</td>
<td>0.9/HO.L+0.1/HO.M</td>
<td>High</td>
</tr>
<tr>
<td><em>Obj</em>(^{(7)})</td>
<td>0.07/NC.M+0.9/NC.H</td>
<td>0.2/BU.M+0.75/BU.H</td>
<td>High</td>
</tr>
<tr>
<td><em>Obj</em>(^{(8)})</td>
<td>0.42/NC.M+0.6/NC.H</td>
<td>0.85/BU.L+0.1/BU.M</td>
<td>Low</td>
</tr>
</tbody>
</table>
Step 3: Assume the threshold value for $\alpha$ - cut is set at 0.4. The linguistic terms in each object with their membership values less than 0.4 are then discarded. The revised fuzzy data set is then shown in Table 4.

Table 4: The fuzzy sets with their membership values equal to or larger than 0.4

<table>
<thead>
<tr>
<th>Object</th>
<th>Transport</th>
<th>Residence</th>
<th>Consumption Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Obj^{(1)}$</td>
<td>0.58/IC.H</td>
<td>0.75/HO.H</td>
<td>High</td>
</tr>
<tr>
<td>$Obj^{(2)}$</td>
<td>1/IC.H</td>
<td>0.83/HO.H</td>
<td>High</td>
</tr>
<tr>
<td>$Obj^{(3)}$</td>
<td>0.5/NC.L</td>
<td>0.85/BU.L</td>
<td>Low</td>
</tr>
<tr>
<td>$Obj^{(4)}$</td>
<td>0.9/NC.L</td>
<td>0.9/BU.L</td>
<td>Low</td>
</tr>
<tr>
<td>$Obj^{(5)}$</td>
<td>0.75/NC.L</td>
<td>0.75/HO.L</td>
<td>Low</td>
</tr>
<tr>
<td>$Obj^{(6)}$</td>
<td>1/NC.L</td>
<td>0.9/HO.L</td>
<td>High</td>
</tr>
<tr>
<td>$Obj^{(7)}$</td>
<td>0.9/NC.H</td>
<td>0.75/BU.H</td>
<td>High</td>
</tr>
<tr>
<td>$Obj^{(8)}$</td>
<td>0.42/NC.M+0.6/NC.H</td>
<td>0.85/BU.L</td>
<td>Low</td>
</tr>
<tr>
<td>$Obj^{(9)}$</td>
<td>1/IC.L</td>
<td>0.75/BU.L</td>
<td>Low</td>
</tr>
<tr>
<td>$Obj^{(10)}$</td>
<td>1/IC.L</td>
<td>0.65/BU.L</td>
<td>Low</td>
</tr>
</tbody>
</table>

Step 4: The terminal-level fuzzy elementary sets of single attributes are found by fuzzy operations based on the definitions in Section 4. The results are shown as follows, with $Transport'$ representing the terminal-level values of the attribute $Transport$.

$$U/\{Transport'\} = \{(\{Obj^{(1)}, Obj^{(2)}\}, 0.58) \cup (\{Obj^{(9)}, Obj^{(10)}\}, 1) \cup (\{Obj^{(3)}, Obj^{(4)}\}, \{Obj^{(5)}, Obj^{(6)}\}, 0.5) \cup (\{Obj^{(7)}, Obj^{(8)}\}, 0.6) \cup (\{Obj^{(8)}\}, 0.42)\};$$

$$U/\{Residence'\} = \{(\{Obj^{(1)}, Obj^{(2)}\}, 0.75) \cup (\{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}, Obj^{(8)}\}, 0.75) \cup (\{Obj^{(3)}, \}.$$
\[ Obj^{(4)}, Obj^{(9)}, Obj^{(10)}, 0.65 \ (\{Obj^{(7)}\}, 0.75). \]

Step 5: \( l \) is set at 1. In this example, assume the class \( X_H \) is first processed.

Step 6: The fuzzy lower approximation of each single terminal-level attribute node for class \( X_H \) is calculated. Take \( \text{Transport} \) as an example to illustrate the step. Since only the terminal-level equivalence class \( \{\text{Obj}^{(1)}, \text{Obj}^{(2)}\}, 0.58 \) for \( \text{Transport} \) are completely include d in \( X_H \), which is \( \{\text{Obj}^{(1)}, \text{Obj}^{(2)}, \text{Obj}^{(7)}, \text{Obj}^{(8)}\}, \) the fuzzy lower approximation of attribute \( \text{Transport} \) for class \( X_H \) is thus:

\[ \text{Transport}^{-}(X_H) = \{(\{\text{Obj}^{(1)}, \text{Obj}^{(2)}\}, 0.58)\}. \]

Similarly, the fuzzy lower approximation of \( \text{Residence} \) for class \( X_H \) is calculated as:

\[ \text{Residence}^{-}(X_H) = \{(\{\text{Obj}^{(1)}, \text{Obj}^{(2)}\}, 0.75) \ (\{\text{Obj}^{(7)}\}, 0.75)\}. \]

Step 7: The fuzzy boundary approximation of each single terminal-level attribute node for class \( X_H \) is calculated. Assume \( \text{Transport} \) is first processed. Since its two equivalence classes \( \{(\text{Obj}^{(3)}, \text{Obj}^{(4)}, \text{Obj}^{(5)}, \text{Obj}^{(6)}\}, 0.5 \) and \( \{(\text{Obj}^{(7)}, \text{Obj}^{(8)}\}, 0.6 \) have the intersection with \( X_H \) and aren’t covered by \( X_H \), the fuzzy boundary approximation of \( \text{Transport} \) for class \( X_H \) is calculated as:

\[ \text{Transport}^{*}(X_H) = \{(\{\text{Obj}^{(3)}, \text{Obj}^{(4)}, \text{Obj}^{(5)}, \text{Obj}^{(6)}\}, 0.5)(\{\text{Obj}^{(7)}, \text{Obj}^{(8)}\}, 0.6)\}. \]

Similarly, the fuzzy boundary approximation of \( \text{Residence} \) for class \( X_H \) is:

\[ \text{Residence}^{*}(X_H) = \{(\{\text{Obj}^{(5)}, \text{Obj}^{(6)}, \text{Obj}^{(8)}\}, 0.75)\}. \]

Step 8: The fuzzy lower and fuzzy boundary approximations of each single non-terminal-level attribute node are calculated from the terminal level to the root.
level by the following substeps.

(a) The fuzzy equivalence class for a non-terminal-level attribute node is derived by the union of the equivalence classes of the same fuzzy regions in its underlying nodes. For example, the equivalence class ($\{Obj^{(1)}, Obj^{(2)}\}, 0.58$) is for the fuzzy region “Imported Car is High,” and ($\{Obj^{(7)}, Obj^{(8)}\}, 0.6$) is for “Native Car is High.” A new equivalence class for “Transport is High” can then be derived by uniting the two equivalence classes as ($\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}, Obj^{(8)}\}, 0.58$).

Here the membership value for the new equivalence class is the minimum of the membership values of the two underlying equivalence classes. Similarly, the equivalence class “Transport is low” is computed as ($\{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(9)}, Obj^{(10)}\}, 0.5$).

The equivalence classes for the other non-terminal-level attribute node $Residence^{nt}$ are calculated as ($\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}\}, 0.75$) and ($\{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(8)}, Obj^{(9)}, Obj^{(10)}\}, 0.65$).

(b) The fuzzy lower approximations of non-terminal-level attribute nodes are calculated. Take the non-terminal-level attribute node $Residence^{nt}$ as an example. One of its equivalence classes, ($\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}\}, 0.75$), is directly put into its fuzzy lower approximation since both its underlying equivalence classes ($\{Obj^{(1)}, Obj^{(2)}\}, 0.75$) and ($\{Obj^{(7)}\}, 0.75$) are lower approximation. Thus:

$$Residence^{nt}(X_H) = \{\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}\}, 0.75\}.$$  

For $Transport^{nt}$ since not all its underlying equivalence classes are in its low-level fuzzy lower approximation, its fuzzy lower approximation is thus empty. Therefore,

$$Transport^{nt}(X_H) = \emptyset.$$
(c) The fuzzy boundary approximations of the non-terminal-level attribute nodes are calculated. Take the attribute node $Transport^{nt}$ for the class $X_H$ as an example. One of its equivalence classes, $\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}, Obj^{(8)}\}$, 0.58, will be put into its non-terminal-level fuzzy boundary approximation for $X_H$ since at least one of its underlying equivalence classes, $\{Obj^{(7)}, Obj^{(8)}\}$, 0.6, is in its lower-level fuzzy boundary approximation. Similarly, its other equivalence class $\{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(9)}, Obj^{(10)}\}$, 0.5 is also put into the non-terminal-level fuzzy boundary approximation. $Transport^{nt}(X_H)$ is thus found as:

$$Transport^{nt}(X_H) = \{\{Obj^{(1)}, Obj^{(2)}, Obj^{(7)}, Obj^{(8)}\}, 0.58\} \cup \{\{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(9)}, Obj^{(10)}\}, 0.5\}.$$ 

Similarly, the non-terminal-level fuzzy boundary approximation for $Residence^{nt}$ is found as:

$$Residence^{nt}(X_H) = \{\{Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}, Obj^{(8)}, Obj^{(9)}, Obj^{(10)}\}, 0.65\}.$$ 

Step 9: $q$ is set at 2, where $q$ is used to count the number of attributes currently being processed.

Step 10: The fuzzy lower and the fuzzy boundary approximations of each attribute set with 2 attributes for class $X_H$ are calculated from the terminal level to the root level. Only a fuzzy region from an attribute can be put in a combination. The combinations from the terminal levels are first processed by the following substeps.

(a) The combinations which have the equivalence class of at least one of its subsets already in the lower approximation for $X_H$ is skipped. In this example, the equivalence classes for the three single attribute values ($Transport = Imported\_Car:High$), ($Residence = Building:High$) and ($Residence = House:High$)
are in the lower approximation for $X_H$. The combinations including these three attribute values won’t then be considered in the later steps. Thus, only the following six combinations for $Transport'$ and $Residence'$ are considered.

$$(Transport=Imported\_Car.\ Low, Residence=House.\ Low),$$

$$(Transport=Imported\_Car.\ Low, Residence=Building.\ Low),$$

$$(Transport=Native\_Car.\ High, Residence=House.\ Low),$$

$$(Transport=Native\_Car.\ High, Residence=Building.\ Low),$$

$$(Transport=Native\_Car.\ Low, Residence=House.\ Low),$$

$$(Transport= Native\_Car.\ Low, Residence=Building.\ Low).$$

(b) The equivalence class of each combination above is found by the intersection of the equivalence classes of its corresponding single attribute regions. Take the combination $(Transport = Native\_Car.Low\; and\; Residence = House.Low)$ as an example. The equivalence class of $(Transport = Native\_Car.Low)$ is $\{(Obj^{(3)}, Obj^{(4)}, Obj^{(5)}, Obj^{(6)}), 0.5\}$ and the one of $(Residence = House.Low)$ is $\{(Obj^{(5)}, Obj^{(6)}), 0.75\}$. The equivalence class of $(Transport = Native\_Car.Low\; and\; Residence = House.Low)$ is thus their intersection, which is $\{(Obj^{(5)}, Obj^{(6)}), 0.75\}$. Note that the membership value is set as the minimum membership values of the two of the subsets. All the equivalence classes for the above combinations of \{Transport', Residence'\} can be similarly derived as follows:

$$U/\{Transport', Residence'\} = \{(Obj^{(3)}, Obj^{(4)}), 0.5\} \cup \{(Obj^{(6)}), 0.6\} \cup \{(Obj^{(5)}), 0.75\} \cup \{(Obj^{(8)}), 0.6\}.$$

(c) The equivalence class of each combination in substep (b) is put into the lower approximation for class $X_H$ if it is covered by $X_H$. Since all the equivalence
classes in the above combinations aren’t covered by the class \(X_H\), the fuzzy lower approximation for the attribute combination \(\{\text{Transport}', \text{Residence}'\}\) is thus empty.

(d) The equivalence class of each combination in substep (b) is put into the boundary approximation class for class \(X_H\) if its intersection with \(X_H\) is not empty. Since the equivalence class of \(\{\text{Transport}', \text{Residence}'\}\), \(\{(\text{Obj}^3, \text{Obj}^4, \text{Obj}^5, \text{Obj}^6), 0.5\}\), contains the object \(\text{Obj}^6\) in class \(X_H\), the fuzzy boundary approximation of \(\{\text{Transport}', \text{Residence}'\}\) is shown below.

\[
\{\text{Transport}', \text{Residence}'\}^*(X_H) = \{(\text{Obj}^5, \text{Obj}^6), 0.75\}.
\]

After the fuzzy lower and the fuzzy boundary approximations for terminal-level attribute combinations are found, the above substeps are repeated for higher-level attribute combinations. They are \(\{\text{Transport}', \text{Residence}''\}\), \(\{\text{Transport}''', \text{Residence}'\}\) and \(\{\text{Transport}''', \text{Residence}'''\}\) in this example. Since the equivalence class for \((\text{Residence} = \text{High})\) is in the lower approximation for \(X_H\), the combinations including \((\text{Residence} = \text{High})\) won’t then be considered. The equivalence classes for the higher-level attribute combinations are found as follows.

\[
\begin{align*}
U/\{\text{Transport}', \text{Residence}''\} &= \{(\text{Obj}^3, \text{Obj}^4, \text{Obj}^5, \text{Obj}^6), 0.5\} (\{\text{Obj}^9, \text{Obj}^{10}\}, 0.65) (\{\text{Obj}^8\}, 0.6); \\
U/\{\text{Transport}''', \text{Residence}'\} &= \{(\text{Obj}^3, \text{Obj}^4, \text{Obj}^9, \text{Obj}^{10}), 0.5\} (\{\text{Obj}^5, \text{Obj}^6\}, 0.75) (\{\text{Obj}^8\}, 0.6); \\
U/\{\text{Transport}''', \text{Residence}'''\} &= \{(\text{Obj}^3, \text{Obj}^4, \text{Obj}^5, \text{Obj}^6, \text{Obj}^9, \text{Obj}^{10}), 0.5\} (\{\text{Obj}^8\}, 0.6). \\
\end{align*}
\]

The fuzzy lower approximations for the above combinations are empty since all the equivalence classes aren’t covered by the class \(X_H\). The fuzzy boundary approximations are found as:

\[
\{\text{Transport}', \text{Residence}''\}^*(X_H) = \{(\text{Obj}^3, \text{Obj}^4, \text{Obj}^5, \text{Obj}^6), 0.5\}.
\]
\[
\{\text{Transport}^*, \text{Residence}^*\}^*(X_H) = \{(\{\text{Obj}^5, \text{Obj}^6\}, 0.75)\};
\]
\[
\{\text{Transport}^*, \text{Residence}^*\}^*(X_H) = \{(\{\text{Obj}^5, \text{Obj}^6\}, 0.75)\} = \{(\{\text{Obj}^3, \\
\text{Obj}^4, \text{Obj}^5, \text{Obj}^6, \text{Obj}^9, \\
\text{Obj}^{10}\}, 0.5)\}.
\]

Step 11: \( q = 2 + 1 = 3 \). Since \( q \) is larger than the number of attributes (= 2), the next step is executed.

Step 12: The linguistic certain rules are derived from the fuzzy lower approximations, and the membership values in the lower approximations are set as the effectiveness measures of the rules for future data. In this example, the following four linguistic certain rules are derived:

1. If \( \text{Transport} \) is \( \text{High\_Cost\ Imported\ Car} \) then \( \text{Consumption\ Style} \) is \( \text{High} \), with future effectiveness = 0.58;
2. If \( \text{Residence} \) is \( \text{High\_Cost\ House} \) then \( \text{Consumption\ Style} \) is \( \text{High} \), with future effectiveness = 0.75;
3. If \( \text{Residence} \) is \( \text{High\_Cost\ Building} \) then \( \text{Consumption\ Style} \) is \( \text{High} \), with future effectiveness = 0.75;
4. If \( \text{Residence} \) is \( \text{High\_Cost} \) then \( \text{Consumption\ Style} \) is \( \text{High} \), with future effectiveness = 0.75.

Step 13: The certain rules more specific than others are removed and the more general ones are reserved. Since the second and third rules are more specific than the fourth rule, the two ones are removed.

Step 14: The linguistic possible rules are derived from the fuzzy boundary approximations, and the membership values in the boundary approximations are set as the effectiveness measures of the rules for future data. The plausibility measure of each rule is also calculated. For example, from \( \text{Transport}^*\ (X_H) = \{(\{\text{Obj}^3, \text{Obj}^4, \text{Obj}^5\} \),
$\text{Obj}^{(6)}$,$0.5$ ($\{\text{Obj}^{(7)}, \text{Obj}^{(8)}\}, 0.6$), the following two possible rules are derived:

1. If Transport is Low_Cost Native Car then Consumption Style is High, with a plausibility $\frac{0.75}{0.5 + 0.9 + 0.75 + 1} = 0.23$, and a future effectiveness $= 0.1$;

2. If Transport is High_Cost Native Car then Consumption Style is High, with a plausibility $\frac{0.9}{0.9 + 0.6} = 0.6$, and a future effectiveness $= 0.1$;

All the other possible rules are derived in the same way. They are shown as follows.

3. If Residence is Low_Cost House then Consumption Style is High, with a plausibility $= 0.35$, and a future effectiveness $= 0.75$;

4. If Transport is High_Cost then Consumption Style is High, with a plausibility $= 0.51$, and a future effectiveness $= 0.58$;

5. If Transport is Low_Cost then Consumption Style is High, with a plausibility $=0.19$, and a future effectiveness $= 0.5$;

6. If Residence is Low_Cost then Consumption Style is High, with a plausibility $=0.15$, and a future effectiveness $= 0.65$;

7. If Transport is Low_Cost Native Car and Residence is Low_Cost House then Consumption Style is High, with a plausibility $= 0.53$, and a future effectiveness $= 0.75$;

8. If Transport is Low_Cost Native Car and Residence is Low_Cost then Consumption Style is High, with a plausibility $= 0.28$, and a future effectiveness $= 0.5$;

9. If Transport is Low_Cost and Residence is Low_Cost House then Consumption Style is High, with a plausibility $= 0.53$, and a future effectiveness $= 0.75$. 

26
effectiveness = 0.75;

10. If Transport is Low_Cost and Residence is Low_Cost then Consumption Style is High, with a plausibility = 0.19, and a future effectiveness = 0.5.

Step 15: Since the tenth rule is more specific than the fifth one and their plausibility values are equal. The former one is then removed from the set of possible rules.

Step 16: \( l = l + 1 = 1 + 1 = 2 \). Steps 6 to 15 are then repeated for the other class \( X_L \).

Step 17: All the linguistic certain rules and possible rules are output. They are shown as follows.

Linguistic certain rules:

1. If Transport is High_Cost Imported_Car then Consumption Style is High, with future effectiveness = 0.58;
2. If Residence is High_Cost then Consumption Style is High, with future effectiveness = 0.75;
3. If Residence is Low_Cost Building then Consumption Style is Low, with future effectiveness = 0.65;
4. If Transport is Low_Cost Imported Car then Consumption Style is Low, with future effectiveness = 1;
5. If Transport is High_Cost Native Car and Residence is Low_Cost House then Consumption Style is Low, with future effectiveness = 0.6;
6. If Transport is Middle_Cost then Consumption Style is Low, with future effectiveness = 0.42.

Linguistic possible rules:

1. If Transport is Low_Cost Native Car then Consumption Style is High, with a plausibility = 0.23, and a future effectiveness = 0.1;
2. If Transport is High_Cost Native Car then Consumption Style is High, with a
3. If *Residence* is *Low_Cost House* then *Consumption Style* is *High*, with a plausibility = 0.35, and a future effectiveness = 0.75;

4. If *Transport* is *High_Cost* then *Consumption Style* is *High*, with a plausibility = 0.51, and a future effectiveness = 0.58;

5. If *Transport* is *Low_Cost* then *Consumption Style* is *High*, with a plausibility = 0.19, and a future effectiveness = 0.5;

6. If *Residence* is *Low_Cost* then *Consumption Style* is *High*, with a plausibility = 0.15, and a future effectiveness = 0.65;

7. If *Transport* is *Low_Cost* *Native Car* and *Residence* is *Low_Cost House* then *Consumption Style* is *High*, with a plausibility = 0.53, and a future effectiveness = 0.75;

8. If *Transport* is *Low_Cost* *Native Car* and *Residence* is *Low_Cost* then *Consumption Style* is *High*, with a plausibility = 0.28, and a future effectiveness = 0.5;

9. If *Transport* is *Low_Cost* and *Residence* is *Low_Cost House* then *Consumption Style* is *High*, with a plausibility = 0.53, and a future effectiveness = 0.75;

10. If *Residence* is *Low_Cost House* then *Consumption Style* is *Low*, with a plausibility = 0.65, and a future effectiveness = 0.75;

11. If *Transport* is *Low_Cost Native Car* then *Consumption Style* is *Low*, with a plausibility = 0.68, and a future effectiveness = 0.5;

12. If *Transport* is *High_Cost Native Car* then *Consumption Style* is *Low*, with a plausibility = 0.4, and a future effectiveness = 0.6;

13. If *Transport* is *Low_Cost* then *Consumption Style* is *Low*, with a plausibility
= 0.58, and a future effectiveness = 0.5;

14. If Residence is Low_Cost then Consumption Style is Low, with a plausibility = 0.15, and a future effectiveness = 0.75;

15. If Transport is High_Cost then Consumption Style is Low, with a plausibility = 0.19, and a future effectiveness = 0.58.

After Step 17, all the linguistic certain and possible rules are derived, and can serve as meta-knowledge concerning the given data set.

7. Discussion

In the proposed learning algorithm for handling training examples with hierarchical values, only the maximally general certain rules, instead of all certain ones, are kept for classification. Certain rules which are not maximally general are removed since they provide no other new information. Take the maximally general rule “If Residence is High_Cost then Consumption Style is High, with future effectiveness = 0.75”, derived in the above section as an example. All the descendent rules covered by the maximally general rule according to the taxonomy relation in Figure 2 are shown as follows:

1. If Residence is High_Cost Building then Consumption Style is High, with future effectiveness = 0.75;

2. If Residence is High_Cost House then Consumption Style is High, with future effectiveness = 0.75.

It can be easily verified that the above two rules are also certain rules. Besides, any rules generated by adding additional constraints into the maximally general rule or into its descendent rules are also certain. These include the following 18 rules:
1. If Transport is Car and Residence is Villa then Consumption Style is High;
2. If Transport is Car and Residence is Single House then Consumption Style is High;
3. If Transport is Car and Residence is Suite then Consumption Style is High;
4. If Transport is Car and Residence is Apartment then Consumption Style is High;
5. If Transport is Car and Residence is House then Consumption Style is High;
6. If Transport is Car and Residence is Building then Consumption Style is High;
7. If Transport is Expensive Car and Residence is Villa then Consumption Style is High;
8. If Transport is Expensive Car and Residence is Single House then Consumption Style is High;
9. If Transport is Expensive Car and Residence is Suite then Consumption Style is High;
10. If Transport is Expensive Car and Residence is Apartment then Consumption Style is High;
11. If Transport is Expensive Car and Residence is House then Consumption Style is High;
12. If Transport is Expensive Car and Residence is Building then Consumption Style is High;
13. If Transport is Cheap Car and Residence is Villa then Consumption Style is High;
14. If Transport is Cheap Car and Residence is Single House then Consumption Style is High;
15. If Transport is Cheap Car and Residence is Suite then Consumption Style is High;
16. If Transport is Cheap Car and Residence is Apartment then Consumption Style is High;
17. If *Transport* is *Cheap Car* and *Residence* is *House* then *Consumption Style* is *High*;

18. If *Transport* is *Cheap Car* and *Residence* is *Building* then *Consumption Style* is *High*;

The pruning procedure is embedded in the proposed algorithm. The above subsumption relation for certain rules is, however, not valid for possible rules. The plausibility of a parent possible rule will always lie between the minimum and the maximum plausibility values of its children rules. Take the possible rule “If *Transport* is *Low_Cost* then *Consumption Style* is *High*, with a plausibility = 0.19” derived in the above section as an example. Both its descendent rules according to the taxonomy relation in Figure 2 are shown as follows:

1. If *Transport* is *Low_Cost Native Car* then *Consumption Style* is *High* with a plausibility = 0.23;

2. If *Transport* is *Low_Cost Imported Car* then *Consumption Style* is *High* with a plausibility = 0.

It can be seen that the plausibility of the parent rule is between 0.23 and 0. Note that the second child rule will not be actually kept since its plausibility is zero. It is only shown here to demonstrate the relationship of the plausibility values in parent and child rules. The child rules with plausibility values less than their parent rules will also be kept by the proposed algorithm since they may provide some useful information about the classification. When a new event satisfies both a child rule and its parent rule, it is more accurate to derive the plausibility of the consequence from the child rule than from the parent rule. However, if a new event has an unknown attribute value, but a known non-terminal value, it can still be
inferred using the parent rules. The proposed algorithm thus keeps all the possible rules except for those with plausibility = 0. If the child rules with plausibility values less than their parent rules won’t be kept, the proposed algorithm can be easily modified by simply adding a subsumption checking after the step of generating the possible rules.

Besides, a plausibility threshold can be used in the proposed algorithm to avoid overwhelming possible rules. The rules with their plausibility values less than the threshold will thus be pruned. This checking step can easily be embedded in finding the boundary approximation to reduce the computational time further.

8. Conclusions and Future Works

In this paper, we have proposed a new learning algorithm based on fuzzy rough sets to find fuzzy cross-level certain and possible rules from training data with hierarchical and quantitative attribute values. The proposed method adopts the concept of fuzzy equivalence classes to find the terminal-level elementary sets of single attributes. These fuzzy equivalence classes are then easily used to find the non-terminal-level elementary sets of single attributes and the cross-level elementary sets of multiple attributes by the union and the intersection operations. Fuzzy lower and fuzzy boundary approximations are then derived from the elementary sets from the terminal level to the root level. Fuzzy boundary approximations, instead of fuzzy upper approximations, are used in the proposed algorithm to find fuzzy possible rules, thus reducing some subsumption checking. Fuzzy lower approximations are used to derive maximally general fuzzy certain rules. Some pruning heuristics are also used to avoid unnecessary search. The fuzzy rules derived can be used to infer results from a new event with both terminal and non-terminal attribute nodes. In the future, we will try to handle other kinds of learning or mining problems.
References


