Learning Maximally General Fuzzy Rules from Rough Sets

Tzung-Pei Hong, Tzu-Ting Wang, Shyue-Liang Wang and Been-Chian Chien

Graduate School of Information Engineering
I-Shou University
Kaohsiung, 84008, Taiwan, R.O.C.
e-mail: {tphong, m873204m, slwang, cbc}@csa500.isu.edu.tw

ABSTRACT

In this paper, we deal with the problem of producing a set of maximally general fuzzy rules from quantitative data. A rule is maximally general if no other rule exists that is more specific than it. We propose a method, which combines the rough set theory and the fuzzy set theory to solve this problem. The proposed method first transforms each quantitative value into a fuzzy set of linguistic terms using membership functions and then calculates the fuzzy lower approximations and the fuzzy upper approximations. The maximally general fuzzy rules are then generated based on these fuzzy approximations by an iterative induction process. The rules derived can then be used to build a prototype knowledge base in a fuzzy system.

Keywords: Fuzzy set, rough set, data mining, certain rule, possible rule.

1. Introduction

Although a wide variety of expert systems have been built, a development bottleneck occurs in knowledge acquisition. Building a large-scale expert system often involves creating and extending a large knowledge base over the course of many months or years [1][20]. Shortening the development time is then the most important factor for the success of expert systems. Machine learning techniques have thus been developed to ease the knowledge-acquisition bottleneck. Among machine learning approaches, deriving inference rules from training examples is the most common [11][12][15][16][22]. Given a set of examples and counterexamples of a concept, the
learning program tries to induce general rules that describe all of the positive training instances and none of the counterexamples (Figure 1). If the training instances belong to more than two classes, the learning program tries to induce general rules for describing each class (Figure 2). Machine learning then provides a feasible way to build a prototype expert system.

Figure 1: Two-classes learning

Figure 2: Three-classes learning

Recently, the rough-set theory has been used in reasoning and knowledge acquisition for expert systems [5][17]. It was proposed by Pawlak in 1982 [18] 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 [17], Germano and Alexandre's knowledge-base reduction [3], Lingras and Yao's data mining [13], Zhong et al.'s rule discovery [25]. Because of the success of the rough-set theory in knowledge acquisition, many researchers in the database and machine-learning fields are very interested in this new research topic since it offers opportunities to discover useful information in training examples. Most previous studies have only shown, however, how binary or crisp valued training data may be handled. Training data in real-world applications sometimes consist of quantitative values, so designing a sophisticated learning algorithm able to deal with various types of data presents a challenge to workers in this research field.
In this paper, we deal with the problem of producing a coverage set of maximally general rules from quantitative data. We propose a method that combines the rough-set theory and the fuzzy-set theory to solve this problem. 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 [4]. They have been applied to many fields such as manufacturing, engineering, diagnosis, and economics [24][26][27]. A new generalized fuzzy learning algorithm based on the rough-set theory is proposed to deal with these fuzzy values and induce maximally general fuzzy rules.

2. Review of the rough-set theory

The rough-set theory, proposed by Pawlak in 1982 [18], 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 $v_j^{(i)}$ 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, $v_j^{(i)} = v_j^{(k)}$), $Obj^{(i)}$ 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:

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

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 upper 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 [5].

3. Notation

The following notation is used in this paper.

\( U \): the universe of all objects;
\( n \): the total number of training examples (objects) in \( U \);
\( Obj^i \): the \( i \)-th training example (object), \( 1 \leq i \leq n \);
\( A \): the set of all attributes describing \( U \);
\( m \): the total number of attributes in \( A \);
\( B \): an arbitrary subset of \( A \);
\( A_j \): the \( j \)-th attribute, \( 1 \leq j \leq m \);
\( |A_j| \): the number of fuzzy regions for \( A_j \);
\( R_{jk} \): the \( k \)-th fuzzy region of \( A_j \), \( 1 \leq k \leq |A_j| \);
\( v_j^{(i)} \): the quantitative value of \( A_j \) for \( Obj^i \);
\( f_j^{(i)} \): the fuzzy set converted from \( v_j^{(i)} \);
\( f_{jk}^{(i)} \): the membership value of \( v_j^{(i)} \) in Region \( R_{jk} \);
\( C \): the set of classes to be determined;
\( c \): the total number of classes in \( C \);
\( x_l \): the \( l \)-th class, \( 1 \leq l \leq c \).

When the same linguistic term \( R_{jk} \) of an attribute \( A_j \) exists in two fuzzy objects \( Obj^0 \) and \( Obj^r \) with membership values \( f_{jk}^{(i)} \) and \( f_{jk}^{(r)} \) larger than zero, \( Obj^0 \) and \( Obj^r \) are said to have a fuzzy indiscernibility relation (or fuzzy 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 \( Obj^0 \) and \( Obj^r \) with membership values larger than zero, \( Obj^0 \) and \( Obj^r \) are said to have a fuzzy indiscernibility relation (or a fuzzy equivalence relation) on attribute subset \( B \) with a membership value equal to the minimum of all the membership values. These fuzzy 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 \( Obj^0 \), is
denoted $B(Obj^i)$. Thus, $B(Obj^i) = \{(B_j(Obj^i), \mu_{B_j}(Obj^i)), \ldots, (B_r(Obj^i), \mu_{B_r}(Obj^i))\}$, 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. The set of fuzzy equivalence classes for a subset set $B$ is referred to as a fuzzy $B$-elementary set.

Fuzzy lower and fuzzy upper 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 upper approximations for $B$ on $X$, denoted $B^-(X)$ and $B^+(X)$, respectively, are defined as follows:

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

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

Elements in $B^-(x)$ 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 upper approximations.

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

4. A fuzzy rough-set algorithm for learning maximally general rules

The details of the proposed fuzzy learning algorithm are described as follows.

Input: A quantitative data set with $n$ objects belonging to $c$ classes, each object having $m$ attribute values, and a set of membership functions.

Output: A set of maximally general 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 $v_{j}^{(i)}$ of each object $Obj^{(i)}$, $i=1$ to $n$, for each
attribute $A_j$, $j=1$ to $m$, into a fuzzy set $f_j^{(i)}$, represented as

$$\left( \frac{f_j^{(i)}}{R_{j_1}^{k_1}} + \frac{f_j^{(i)}}{R_{j_2}^{k_2}} + \ldots + \frac{f_j^{(i)}}{R_{j_l}^{k_l}} \right),$$

using the given membership functions, where $R_{jk}$ is the $k$-th fuzzy region of attribute $A_j$, $f_j^{(i)}$ is $v_j^{(i)}$’s fuzzy membership value in region $R_{jk}$, and $l=|A_j|$ is the number of fuzzy regions for $A_j$.

Step 3: Find the fuzzy elementary sets of the singleton attributes using fuzzy operations

Step 4: Initialize $l = 0$, where $l$ is used to count the number of the class being processed and $k$ is used to count the number of attributes being processed.

Step 5: Set $l = l + 1$.

Step 6: Set $q = 0$, where $q$ is used to count the number of attributes being processed.

Step 7: Set $q = q + 1$.

Step 8: Compute the fuzzy lower approximation of each subset $B$ with $q$ attributes for class $X_l$ as:

$$B^*(X_l) = \{(B_k(x), \mu_{B_k}(x)) \mid x \in U, B_k(x) \subseteq X_l, 1 \leq k \leq |B(x)|\}.$$

Step 9: Choose the $B'_k(x)$ with the maximum number of $|B'_k(x)|$; put it in the certain rule set.

Step 10: Remove the elements in $B'_k(x)$ from any $B_k(x)$ and from $X_l$.

Step 11: Repeat Steps 9 and 10 until all $B_k(x)$’s are empty.

Step 12: If $X_l$ is empty, go to Step 20; if $X_l$ is not empty and $q \neq m$, go to Step 7; otherwise, do the next step.

Step 13: Compute the fuzzy upper approximation of each singleton $B$ (with 1 attribute) for class $X_l$ as:

$$B^*(X_l) = \{(B_k(x), \mu_{B_k}(x)) \mid x \in U, B_k(x) \cap X_l \neq \emptyset, 1 \leq k \leq |B(x)|\}.$$

Step 14: Calculate the plausibility measure of each $B_k(x)$ as:

$$p(B_k(x)) = \frac{\sum_{x \in B_k(x) \cap \text{original } X_l} \mu_{B_k}(x)}{\sum_{x \in B_k(x)} \mu_{B_k}(x)},$$

where original $X_l$ means the object set of class $l$ in Step 2.

Step 15: Choose the subset $B'_k(x)$ with the maximum plausibility measure; put it in the possible rule set.

Step 16: Set $X'_l = X_l$ and Remove the elements in $B'_k(x)$ from any $B_k(x)$ and from $X_l$. 
Step 17: If $B_k(x)$ contains no elements in $X_l$, remove it from the fuzzy upper approximation.

Step 18: Recursively check whether $B'_k(x)$'s specialization can generate a set of new $B''_k(x)$ covering the elements in $(B'_k(x) \cap X'_l)$ with plausibility measure larger than that of $B'_k(x)$. Replace $B'_k(x)$ with $B''_k(x)$ in the possible rule set.

Step 19: Repeat Steps 15 to 18 until $X_l$ is empty.

Step 20: Repeat Steps 5 to 19 until $l = c$.

Step 21: Derive the certain rules from the certain rule set.

Step 22: Derive the possible rules from the possible rule set.

After Step 22, maximally general certain and possible rules can be derived, and can serve as meta-knowledge concerning the given data set.

5. **An example**

In this section, an example is given to show how the proposed algorithm can be used to generate maximally general certain and possible rules from quantitative data. Table 1 shows a quantitative data set.

<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>122</td>
<td>80</td>
<td>N</td>
</tr>
<tr>
<td>Obj(2)</td>
<td>155</td>
<td>90</td>
<td>H</td>
</tr>
<tr>
<td>Obj(3)</td>
<td>130</td>
<td>92</td>
<td>N</td>
</tr>
<tr>
<td>Obj(4)</td>
<td>87</td>
<td>68</td>
<td>L</td>
</tr>
<tr>
<td>Obj(5)</td>
<td>165</td>
<td>93</td>
<td>H</td>
</tr>
<tr>
<td>Obj(6)</td>
<td>139</td>
<td>100</td>
<td>H</td>
</tr>
<tr>
<td>Obj(7)</td>
<td>95</td>
<td>75</td>
<td>L</td>
</tr>
</tbody>
</table>

Assume the membership functions for each attribute are given by experts as shown in Figure 3.

![Figure 3: The given membership functions of each attribute](image)

Table 1: A quantitative data set as an example
The quantitative values of each object are transformed into fuzzy sets. Results for all the objects are shown in Table 2.

Table 2: The fuzzy sets transformed from the data in Table 1.

<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>(0.9/N)</td>
<td>(0.9/N)</td>
<td>N</td>
</tr>
<tr>
<td>Obj^{(2)}</td>
<td>(0.1/N + 0.75/H)</td>
<td>(0.4/N)</td>
<td>H</td>
</tr>
<tr>
<td>Obj^{(3)}</td>
<td>(0.85/N)</td>
<td>(0.3/N + 0.4/H)</td>
<td>N</td>
</tr>
<tr>
<td>Obj^{(4)}</td>
<td>(1/L)</td>
<td>(1/L)</td>
<td>L</td>
</tr>
<tr>
<td>Obj^{(5)}</td>
<td>(1/H)</td>
<td>(0.16/N + 0.6/H)</td>
<td>H</td>
</tr>
<tr>
<td>Obj^{(6)}</td>
<td>(0.4/N)</td>
<td>(1/H)</td>
<td>H</td>
</tr>
<tr>
<td>Obj^{(7)}</td>
<td>(0.5/L + 0.1/N)</td>
<td>(0.4/N)</td>
<td>L</td>
</tr>
</tbody>
</table>

The fuzzy elementary sets of the singleton attributes \( SP \) and \( DP \) are found as follows:

\[
U/\{SP\} = \{(\{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}, Obj^{(6)}, Obj^{(7)}\}, 0.1) (\{Obj^{(2)}, Obj^{(5)}\}, 0.75) (\{Obj^{(4)}, Obj^{(7)}\}, 0.5)\},
\]

\[
U/\{DP\} = \{(\{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}, Obj^{(5)}, Obj^{(7)}\}, 0.3) (\{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\}, 0.4) (\{Obj^{(4)}\}, 1)\}.
\]

The fuzzy lower approximation of single attribute \( q=1 \) for class \( X_H \) is first calculated. Since only \( \{Obj^{(2)}, Obj^{(5)}, Obj^{(6)}\} \) is included in \( X_H \), thus:

\[
SP^*(X_H) = \{(\{Obj^{(2)}, Obj^{(5)}\}, 0.75)\},
\]

\[
DP^*(X_H) = \emptyset.
\]

The subset \( \{Obj^{(2)}, Obj^{(5)}\} \) is first put in the certain rule set. The elements in \( \{Obj^{(2)}, Obj^{(5)}\} \) are then removed from \( SP^*(X_H), DP^*(X_H) \) and \( X_H \). Thus:

\[
SP^*(X_H) = \emptyset, DP^*(X_H) = \emptyset, \text{ and } X_H = \{Obj^{(6)}\}.
\]

The fuzzy upper approximation of each singleton attribute for class \( X_H \) is calculated. Since only \( \{Obj^{(6)}\} \) is included in \( X_H \) after Step 12, thus:

\[
SP^*(X_H) = \{(\{Obj^{(1)}, Obj^{(2)}, Obj^{(6)}, Obj^{(7)}\}, 0.1)\}, \text{ and }
\]

\[
DP^*(X_H) = \{(\{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\}, 0.4)\}.
\]

Since \( DP_H(x) = \{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\} \) has the maximum plausibility measure, it is put in the possible rule set. \( DP_H(x) = \{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\} \) is then specialized by the attribute \( SP \). Since none of its specialization has a larger plausibility measure than it, no replacement is done. Similar steps can be easily derived for finding the certain and possible rule sets for classes \( X_N \) and \( X_L \).

### 6. Conclusion

In this paper, we have proposed a generalized data-mining algorithm that can process data with quantitative values. The algorithm integrates the fuzzy-set and the
rough-set concepts, and discovers maximally-general certain and possible rules. The rules thus mined exhibit fuzzy quantitative regularity in databases and can be used to provide suggestions to appropriate supervisors.

References


