PERGAMON

Expert Systems with Applications 19 (2000) 97–103

Expert Systems with Applications

www.elsevier.com/locate/eswa

Learning a coverage set of maximally general fuzzy rules by rough sets

T.-P. Hong*, T.-T. Wang, S.-L. Wang, B.-C. Chien

Institute of Information Engineering, I-Shou University, Kaohsiung 84008, Taiwan, ROC

Abstract

Expert systems have been widely used in domains where mathematical models cannot be easily built, human experts are not available or the cost of querying an expert is high. Machine learning or data mining can extract desirable knowledge or interesting patterns from existing databases and ease the development bottleneck in building expert systems. In the past we proposed a method [Hong, T.P., Wang, T.T., Wang, S.L. (2000). Knowledge acquisition from quantitative data using the rough-set theory. *Intelligent Data Analysis* (in press).], which combined the rough set theory and the fuzzy set theory to produce all possible fuzzy rules from quantitative data. In this paper, we propose a new algorithm to deal with the problem of producing a set of maximally general fuzzy rules for coverage of training examples from quantitative data. A rule is maximally general if no other rule exists that is both more general and with larger confidence than it. 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 expert system.

Keywords: Machine learning; Fuzzy set; Rough set; Data mining; Expert system

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 (Buchanan & Shortliffe, 1984; Riley, 1989). 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 (Kodratoff & Michalski, 1983; Krone & Kiendl, 1997; Michalski, Carbonell & Mitchell, 1983a, 1983b; Michalski, Carbonell & Mitchell, 1984; Tsumoto, 1998). 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 (Fig. 1). If the training instances belong to more than two classes, the learning program tries to induce general rules for describing each class (Fig. 2). Machine learning then provides a feasible way to build a prototype expert system.

Recently, the rough-set theory has been used in reasoning and knowledge acquisition for expert systems (Grzymala-Busse, 1988; Orlowska, 1990). It was proposed by Pawlak (1982)) 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 (1990) reasoning with incomplete information, Germano and Alexandre's (1996) knowledge-base reduction, Lingras and Yao's (1998) data mining, Zhong, Dong and Ohsuga's (1998) rule discovery. 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.

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 (Graham & Jones, 1988). They have been applied to many fields such as manufacturing, engineering, diagnosis, and economics (Zadeh, 1988; Zimmermann, 1987; Zimmermann, 1991). In the

* Corresponding author. Tel.: +886-7-6563711, ext. 6569; fax: +886-7-6563734.

E-mail address: tphong@isu.edu.tw (T.-P. Hong).

0957-4174/00/$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.
PII: S0957-4174(00)00024-5

Nomenclature

U	the universe of all objects
N	the total number of training examples (objects) in U
$Obj^{(i)}$	the ith training example (object), $1 \le i \le n$
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 jth attribute, $1 \le j \le m$
$\lvert A_j \rvert$	the number of fuzzy regions for A_j
R_{jk}	the kth fuzzy region of A_j, $1 \le k \le \lvert A_j \rvert$
$v_j^{(i)}$	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 lth class, $1 \le l \le c$.

past, we proposed a method (Hong, Wang & Wang, 2000), which combined the rough set theory and the fuzzy set theory to produce all possible fuzzy rules from quantitative data. In this paper, we propose a new algorithm to deal with the problem of producing a set of maximally general fuzzy rules for coverage of training examples from quantitative data. A rule is maximally general if no other rule exists that is both more general and with larger confidence than it (Hong & Tseng, 1997). 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 remainder of this paper is organized as follows. In Section 2, the integration of rough sets and fuzzy sets is briefly reviewed. In Section 3, a new learning algorithm based on the rough-set theory and the fuzzy-set theory is proposed to induce maximally general fuzzy rules from quantitative values. An example is given to illustrate the proposed algorithm in Section 4. Finally, conclusions are given in Section 5.

2. Integration of fuzzy sets and rough sets

In Hong et al. (2000), we integrated fuzzy sets and rough sets to process quantitative data. The concepts are reviewed as follows.

When the same linguistic term R_{jk} of an attribute A_j exists in two fuzzy objects $Obj^{(i)}$ and $Obj^{(r)}$ with membership values $f_{jk}^{(i)}$ and $f_{jk}^{(r)}$ larger than zero, $Obj^{(i)}$ 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 $Obj^{(i)}$ and $Obj^{(r)}$ with membership values larger than zero, $Obj^{(i)}$ 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 $Obj^{(i)}$, 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 jth partition in $B(Obj^{(i)})$, and $\mu_{B_j}(Obj^{(i)})$ is the membership value of the jth partition.

Example 1. Table 1 shows a data set containing three fuzzy objects $U = \{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}\}$, two attributes $A = \{$*Systolic Pressure* (*SP*), *Diastolic Pressure* (*DP*)$\}$, and a class set *Blood Pressure*(*BP*). The attributes and the class set each have three possible terms: $\{$*Low*(*L*), *Normal*(*N*), *High*(*H*)$\}$. $Obj^{(1)}$ 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 $Obj^{(3)}$ are classified similarly.

Since the same linguistic term (N) for attribute SP exists in both $Obj^{(1)}$ and $Obj^{(3)}$, 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 been formed and may be

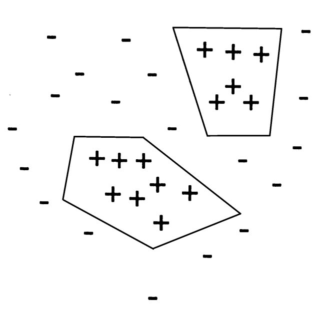

Fig. 1. Two-classes learning.

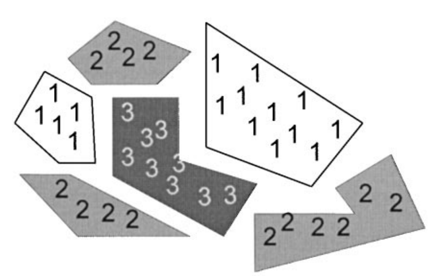

Fig. 2. Three-classes learning.

Table 1
The three fuzzy objects used in Example 1

Object	Systolic pressure (*SP*)	Diastolic pressure (*DP*)	Blood pressure (*BP*)
$Obj^{(1)}$	$(0.1/N + 0.75/H)$	$(0.4/N + 0.8/H)$	H
$Obj^{(2)}$	$(1/H)$	$(0.16/N + 0.6/H)$	H
$Obj^{(3)}$	$(0.5/L + 0.3/N)$	$(0.4/N + 0.3/L)$	L

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 easily be 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 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:

$$B_*(X) = \{(B_k(x), \mu_{B_k}(x)) | x \in U, B_k(x) \subseteq X,\ 1 \le k \le |B(x)|\},$$

and

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

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.

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

$$SP_*(X) = (\{Obj^{(1)}, Obj^{(2)}\}, 0.75),$$

and

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

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. Also, the traditional lower and upper approximations are specialization of the proposed fuzzy lower and upper approximations. For a rough-set algorithm to process crisp data, each attribute value must be crisp such that it can be represented as a linguistic term with a sole membership value = 1, and all other linguistic membership values = 0. In this environment, the traditional lower and upper approximations can be proven equivalent to the fuzzy lower and upper approximations (Hong et al., 2000).

3. Learning maximally general fuzzy rules

In this section, we propose a fuzzy learning algorithm, based on rough sets, for inducing a coverage set of maximally general fuzzy rules from quantitative data. The proposed fuzzy learning algorithm first transforms each quantitative value into a fuzzy set of linguistic terms using membership functions (Hong et al., 2000). The algorithm then calculates the fuzzy lower approximations and the fuzzy upper approximations using fuzzy operations. The rule-derivation process based on these fuzzy approximations is then performed to find maximally general rules in an iterative way. The details of the proposed fuzzy learning algorithm are described as follows.

The fuzzy rough-set algorithm for learning maximally general rules:

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_l as X_l.
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

$$\left(\frac{f_{j_1}^{(i)}}{R_{j_1}} + \frac{f_{j_2}^{(i)}}{R_{j_2}} + \cdots + \frac{f_{j_l}^{(i)}}{R_{j_l}}\right),$$

using the given membership functions, where R_{jk} is the

Table 2
A quantitative data set as an example

Object	Systolic pressure (*SP*)	Diastolic pressure (*DP*)	Blood pressure (*BP*)
$Obj^{(1)}$	122	80	*N*
$Obj^{(2)}$	155	90	*H*
$Obj^{(3)}$	130	92	*N*
$Obj^{(4)}$	87	68	*L*
$Obj^{(5)}$	165	93	*H*
$Obj^{(6)}$	139	100	*H*
$Obj^{(7)}$	95	75	*L*

kth fuzzy region of attribute A_j, $f_{jk}^{(i)}$ is $v_j^{(i)}$'s fuzzy membership value in the 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:

$$B_*(X_l) = \{(B_k(x), \mu_{B_k}(x)) | x \in U, B_k(x) \subseteq X_{l,}\ 1 \le k \le |B(x)|\}.$$

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:

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

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 15–18 until X_l is empty.
Step 20: Repeat Steps 5–19 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.

4. 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 2 shows a quantitative data set which is similar to that shown in Table 1 except that the data attributes are represented as quantitative values.

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

The proposed learning algorithm processes this data set as follows.

Step 1: Since three classes exist in the data set, three partitions are formed as follows:

$$X_H = \{Obj^{(2)}, Obj^{(5)}, Obj^{(6)}\},$$

$$X_N = \{Obj^{(1)}, Obj^{(3)}\},$$

and

$$X_L = \{Obj^{(4)}, Obj^{(7)}\}.$$

Step 2: The quantitative values of each object are transformed into fuzzy sets. Take the attribute *Systolic Pressure* (*SP*) in $Obj^{(2)}$ as an example. The value "155" is converted into a fuzzy set $(0.1/N + 0.75/H)$ using the given membership functions. Results for all the objects are shown in Table 3.
Step 3: The fuzzy elementary sets of the singleton attributes *SP* and *DP* are found as follows:

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

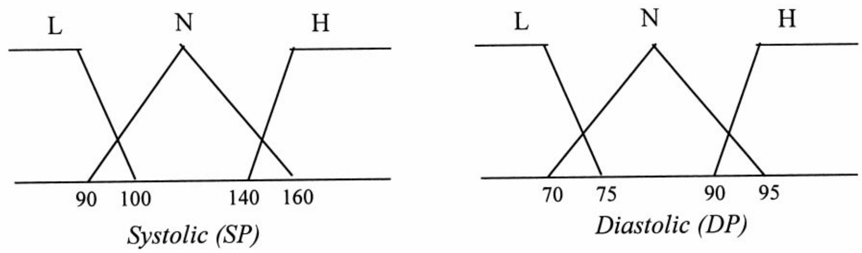

Fig. 3. The given membership functions of each attribute.

and

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

Step 4: l is set to 0.
Step 5: $l = l + 1 = 1$, where l is used to count the identification number of the class being processed.
Step 6: q is set to 0.
Step 7: $q = q + 1 = 1$, where q is used to count the number of attributes being processed.
Step 8: 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:

$$SP_*(X_H) = \{(\{Obj^{(2)}, Obj^{(5)}\}, 0.75)\},$$

$$DP_*(X_H) = \emptyset.$$

Step 9: Since the subset $\{Obj^{(2)}, Obj^{(5)}\}$ has the maximum number of elements in $SP_*(X_H)$ and $DP_*(X_H)$, it is first put in the certain rule set.
Step 10: The elements in $\{Obj^{(2)}, Obj^{(5)}\}$ are then removed from $SP_*(X_H)$, $DP_*(X_H)$ and X_H. Thus:

$$SP_*(X_H) = \emptyset,\ \ DP_*(X_H) = \emptyset,\ \text{ and }\ X_H = \{Obj^{(6)}\}.$$

Step 11: Steps 9 and 10 are repeated until all fuzzy lower approximations are empty. In this case, since both $SP_*(X_H)$ and $DP_*(X_H)$ are empty, the next step is executed.
Step 12: Since X_H is still not empty and $q(1) \neq m(2)$, Steps 7–11 are executed again. $q = 1 + 1 = 2$. The fuzzy lower approximation of two attributes ($q = 2$) for class X_H ($\{Obj^{(6)}\}$) is calculated as:

$$SP, DP_*(X_H) = \emptyset.$$

No subset of attributes is further put in the certain rule set. Since X_H is still not empty and $q(2) = m(2)$, Step 13 is then executed to find possible rules.
Step 13: 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:

$$SP^*(X_H) = \{(\{Obj^{(1)}, Obj^{(3)}, Obj^{(6)}, Obj^{(7)}\}, 0.1)\},$$

and

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

Step 14: The plausibility measure of the subset $SP_N(x)\times$ $(= \{Obj^{(1)}, Obj^{(3)}, Obj^{(6)}, Obj^{(7)}\})$ is calculated as:

$$p(SP_N(x)) = \frac{0.4}{0.9 + 0.85 + 0.4 + 0.1} = 0.18;$$

Similarly, the plausibility measure of the subset $DP_H(x)\times$ $(= \{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\})$ is calculated as:

$$p(DP_H(x)) = \frac{1}{1 + 0.4} = 0.71.$$

Step 15: Since $DP_H(x)$ $(= \{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\})$ has the maximum plausibility measure, it is put in the possible rule set.
Step 16: $X'_H = X_H = \{Obj^{(6)}\}$; $Obj^{(3)}, Obj^{(5)}$ and $Obj^{(6)}$

Table 3
The fuzzy sets transformed from the data in Table 2

Object	Systolic pressure (*SP*)	Diastolic pressure (*DP*)	Blood pressure (*BP*)
$Obj^{(1)}$	(0.9/*N*)	(0.9/*N*)	*N*
$Obj^{(2)}$	(0.1/*N* + 0.75/*H*)	(0.4/*N*)	*H*
$Obj^{(3)}$	(0.85/*N*)	(0.3/*N* + 0.4/*H*)	*N*
$Obj^{(4)}$	(1/*L*)	(1/*L*)	*L*
$Obj^{(5)}$	(1/*H*)	(0.16/*N* + 0.6/*H*)	*H*
$Obj^{(6)}$	(0.4/*N*)	(1/*H*)	*H*
$Obj^{(7)}$	(0.5/*L* + 0.1/*N*)	(0.4/*N*)	*L*

are removed from $SP^*(X_H)$, $DP^*(X_H)$ and X_H. Thus:

$$SP^*(X_H) = \{(\{Obj^{(1)}, Obj^{(7)}\}, 0.1)\},$$

$$DP^*(X_H) = \emptyset,$$

and

$$X_H = \emptyset.$$

Step 17: Since $\{Obj^{(1)}, Obj^{(7)}\}$ contains no elements in X_H, it is removed from the upper approximation. Thus:

$$SP^*(X_H) = \emptyset,$$

$$DP^*(X_H) = \emptyset,$$

and

$$X_H = \emptyset.$$

Step 18: $DP_H(x) \cap X'_H = \{Obj^{(6)}\}$. $DP_H(x)$ $(= \{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\})$ is then specialized by the attribute SP as:

$$DP_H, SP_{\mathrm{L}}(Obj^{(6)}) = \emptyset,$$

$$DP_H, SP_{\mathrm{N}}(Obj^{(6)}) = (\{Obj^{(3)}, Obj^{(6)}\}, 0.4),$$

and

$$DP_H, SP_{\mathrm{H}}(Obj^{(6)}) = \emptyset.$$

The plausibility measure of the subset DP_H,$SP_{\mathrm{N}}(Obj^{(6)})$ is calculated as:

$$p(DP_H, SP_{\mathrm{N}}(Obj^{(6)})) = \frac{0.3}{0.4 + 0.3} = 0.43.$$

Since $p(DP_H, SP_{\mathrm{N}}(Obj^{(6)})) < p(DP_H(Obj^{(6)}))$, (DP_H, SP_{N}) does not replace DP_H in the possible rule set.
Step 19: Since X_H is empty, Step 20 is executed.
Step 20: Steps 5–19 are repeated for finding the certain and possible rule sets for classes X_N and X_L. The final results are shown as follows:
Certain rule sets:

$$X_H = \{Obj^{(2)}, Obj^{(5)}\},$$

$$X_N = \emptyset,$$

and

$$X_L = \{Obj^{(4)}, Obj^{(7)}\}.$$

Possible rule sets:

$$X_H = (\{Obj^{(3)}, Obj^{(5)}, Obj^{(6)}\}, 0.4),$$

$$X_N = (\{Obj^{(1)}, Obj^{(2)}, Obj^{(3)}, Obj^{(7)}\}, 0.1)$$

and

$$X_L = \emptyset.$$

Step 21: The certain rules derived from the certain rule set are shown as follows:
If *Systolic Pressure* = *High* Then *Blood Pressure* = *High*.
If *Systolic Pressure* = *Low* Then *Blood Pressure* = *Low*.
Step 22: The possible rules derived from the possible rule set are shown as follows:
If *Diastolic Pressure* = *High* Then *Blood Pressure* = *High*, with plausibility = 0.71.
If *Systolic Pressure* = *Normal* and *Diastolic Pressure* = *Normal* Then *Blood Pressure* = *Normal*, with plausibility = 0.85.

These maximally general certain and possible rules can then serve as meta-knowledge concerning the given data set. Note that in the above example, only four rules are derived to cover the given training examples. Seventeen rules are however found using the method in (Hong et al., 2000). The proposed method can thus find a concise set of fuzzy rules to cover the given training examples.

5. Conclusion and future work

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 a coverage set of maximally general certain and possible rules. The rules thus learned exhibit fuzzy quantitative regularity in training examples and can be used to provide suggestions to appropriate system designers.

Although the proposed method works well in knowledge acquisition for quantitative values, it is just a beginning. There is still much work to be done in this field. Our method assumes that the membership functions are known in advance. In Hong and Lee (1996), Hong and Tseng (1997), Hong et al. (2000), and Wang, Hong and Tseng (1998), we also proposed fuzzy learning methods for automatically deriving membership functions. In the future, we will attempt to dynamically adjust the membership functions in the proposed algorithm to avoid the membership function acquisition bottleneck.

References

Buchanan, B. G., & Shortliffe, E. H. (1984). *Rule-based expert system: the MYCIN experiments of the Standford Heuristic Programming Projects*, Reading, MA: Addison-Wesley.

Germano, L. T., & Alexandre, P. (1996). Knowledge-base reduction based on rough set techniques. *Canadian Conference on Electrical and Computer Engineering*, 278–281.

Graham, I., & Jones, P. L. (1988). *Expert systems-knowledge, uncertainty and decision* (pp. 117–158). Boston: Chapman and Computing.

Grzymala-Busse, J. W. (1988). Knowledge acquisition under uncertainty: a rough set approach. *Journal of Intelligent Robotic Systems*, *1*, 3–16.

Hong, T. P., & Lee, C. Y. (1996). Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets and Systems*, *84*, 33–47.

Hong, T. P., & Tseng, S. S. (1997). A generalized version space learning algorithm for noisy and uncertain data. *IEEE Transactions on Knowledge and Data Engineering*, *9* (2), 336–340.

Hong, T. P., Kuo, C. S., & Chi, S. C. (2000). Mining association rules from quantitative data. *Intelligent Data Analysis: an International Journal*, *3* (5), 363–376.

Hong, T. P., Wang, T. T., & Wang, S. L. (2000). Knowledge acquisition from quantitative data using the rough-set theory. *Intelligent Data Analysis*, (in press).

Kodratoff, Y., & Michalski, R. S. (1983). *Machine learning: an artificial intelligence artificial intelligence approach 3*, San Mateo, CA: Morgan Kaufmann.

Krone, A., & Kiendl, H. (1997). An evolutionary concept for generating relevant fuzzy rules from data. *International Journal of Knowledge-Based Intelligent Engineering Systems*, *1* (4), 207–213.

Lingras, P. J., & Yao, Y. Y. (1998). Data mining using extensions of the rough set model. *Journal of the American Society for Information Science*, *49* (5), 415–422.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983a). *Machine learning: an artificial intelligence approach 1*, Los Altos, CA: Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1983b). *Machine learning: an artificial intelligence approach*, vol. 1. Palo Alto, CA: Tioga.

Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (1984). *Machine learning: an artificial intelligence approach*, vol. 2. Palo Alto, CA: Tioga.

Orlowska, E. (1994). Reasoning with incomplete information: rough set based information logics. In V. Alagar, S. Bergler & F. Q. Dong, *Incompleteness and uncertainty in information systems* (pp. 16–33).

Pawlak, Z. (1982). Rough set. *International Journal of Computer and Information Sciences*, 341–356.

Riley, G. (1989). *Expert systems—principles and programming*, Boston: Pws-Kent (pp. 1–56).

Tsumoto, S. (1998). Extraction of experts decision rules from clinical databases using rough set model. *Intelligent Data Analysis*, *2*, 215–227.

Wang, C. H., Hong, T. P., & Tseng, S. S. (1998). Integrating fuzzy knowledge by genetic algorithms. *IEEE Transactions on Evolutionary Computation*, *2* (4), 138–149.

Zadeh, L. A. (1988). Fuzzy logic. *IEEE Computer*, 83–93.

Zhong, N., Dong, J. Z., Ohsuga, S., & Lin, T. Y. (1998). An incremental, probabilistic rough set approach to rule discovery. *IEEE International Conference on Fuzzy Systems*, *2*, 933–938.

Zimmermann, H. J. (1987). *Fuzzy sets, decision making, and expert systems*, Boston: Kluwer.

Zimmermann, H. J. (1991). *Fuzzy set theory and its applications*, Boston: Kluwer.